Epigenomic Approaches to Neuropsychiatric Health and Disease Across the Life-course

Marzi, Sarah Julia

Awarding institution:
King's College London

Epigenomic Approaches to Neuropsychiatric Health and Disease Across the Life-course

Sarah Julia Marzi

A thesis submitted to King's College London for the degree of Doctor of Philosophy (PhD) 2017

MRC Social, Genetic and Developmental Psychiatry Centre,
Institute of Psychiatry, Psychology and Neuroscience,
King's College London

Thesis abstract

Neuropsychiatric disorders represent an immense global health and economic burden, and identifying genetic and environmental factors contributing to these diseases is a key challenge in contemporary biomedical research. Efforts to identify genetic variation associated with disease risk have had some notable success, however the proportion of variance in disease explained by known genetic risk variants is limited. There is increasing evidence for epigenetic dysregulation in neuropsychiatric disease, with regulatory variation potentially induced by psychosocial and environmental risk factors associated with pathology. My PhD explores how the analysis of epigenetic variation, using novel technologies and unique study designs, can contribute to our understanding of how genetic and environmental factors influence disease risk.

The first part of my thesis focuses on epigenetic correlates of early-life adversity, an exposure robustly associated with psychiatric phenotypes. Two complementary epidemiological study designs were used to characterise these associations: a “natural experiment”, using a sample of Romanian adoptees (**Chapter 2**), and a large longitudinal population-based twin cohort (**Chapter 3**). I identified a differentially methylated region annotated to *CYP2E1* associated with institutional deprivation and socio-cognitive abilities, but found no robust associations between early-life victimisation and DNA methylation in blood. Neither study identified epigenetic variation associated with early-life stress in previously reported HPA-axis genes.

The second part of my thesis explores epigenetic variation in the ageing brain and neurodegenerative disease. I undertook a systematic characterisation of allele-specific DNA methylation across multiple human brain regions and blood, finding widespread and tissue-specific patterns of allelic-skewing (**Chapter 4**). To further understand epigenetic variation in neurodegenerative disease, I investigated the relationship between the histone modification H3K27ac (which marks active gene expression) and Alzheimer’s disease progression (**Chapter 5**). I found widespread acetylomic dysregulation associated with neuropathology, enriched in disease-specific biological pathways and mapping to regions harbouring genetic variants associated with familial and sporadic AD.

The work presented in this thesis was funded by an Early Stage Researcher Fellowship as part of the Marie Curie Initial Training Network “EpiTrain” (REA grant agreement no. 316758).

Approximate word count: 60,000

Declarations

Chapter 2

Sample and data collection was completed by investigators of the English and Romanian Adoption Study. Laboratory procedures were undertaken by Dr Joana Viana, Dr Emma Dempster and Bethany Crawford. I undertook all analyses, and drafted the manuscript in collaboration with Professor Jonathan Mill, Professor Robert Kumsta and Professor Edmund Sonuga-Barke. Professor Kumsta and I shared joint first-authorship of the paper.

Chapter 3

Sample and data collection for the E-Risk Longitudinal Twin Study was completed by the E-Risk Study team. Laboratory procedures were performed by Dr Joe Burrage and data pre-processing by Dr Eilis Hannon. The cumulative victimisation measure was derived by Professor Avshalom Caspi. I undertook all data analysis and drafted the manuscript text.

Chapter 4

Post-mortem brain and pre-mortem whole blood samples were provided by the MRC London Neurodegenerative Disease Brain Bank. Laboratory work was performed by Dr Jose Paya-Cano. Data pre-processing was performed by Professor Leonard Schalkwyk. All data analyses were conducted by myself. I wrote the manuscript in collaboration with Professor Jonathan Mill, with editorial input from Professor Leonard Schalkwyk and Dr Emma Meaburn.

Chapter 5

Post-mortem brain samples were provided by the MRC London Neurodegenerative Disease Brain Bank. Chromatin immunoprecipitation and sequencing library preparation was performed by myself in collaboration with Dr Teodora Ribarska. The libraries were sequenced at the University of Exeter Sequencing Service. All data pre-processing and analyses were conducted by

myself. Linkage disequilibrium blocks for the GWAS overlap analysis were generated by Dr Eilis Hannon. Pre-processed DNA methylation and hydroxymethylation data were provided by Dr Katie Lunnon and Adam Smith.

Acknowledgements

The immense effort that goes into a PhD over several years is never possible without a good research and social network and I have had a large and fantastic network of people around me to thank for.

First, I would like to thank my parents for their continued support, patience, encouragement and love throughout my academic journey. None of this would have been possible without you, obviously.

A massive thanks to my three fantastic supervisors: Jon, it has been an absolute pleasure working with you. Thanks so much for all your guidance, brilliance, dedication and for giving me the opportunity to explore these fascinating scientific avenues. Leo, thanks equally for your support, genius inputs, computational magic and encouragement. Chloe, many thanks for all the personal support, advice, chats and keeping the Psychiatric Epigenetics group spirits high here at the centre. Huge thanks also to Stuart, whose informatics wizardry resolved even my most mysterious computational and software problems on any cluster.

I would like to thank all the current and former members of the Complex Disease Epigenetics group as well as my PhD cohort and office mates. You have been the best colleagues, PhD companions, friends, coffee/lunch club associates I could have hoped for. I am also grateful to amazing collaborators on all of my projects, especially the teams behind the English and Romanian Adoption Study, the E-Risk Longitudinal Twin Study and colleagues from my Marie Curie Network "EpiTrain", who hosted me at their labs, trained me and worked with me. It's been a wonderful experience!

Finally, Niklas, you've been my rock, PhD accomplice, grounding influence and inspiration. I am so grateful we embarked on this journey (to the UK and through the PhD) together. I'm not sure I would have commenced or completed it without you – thanks so much for everything.

Publications arising from this thesis

Chapter 2

Kumsta, R.*, **Marzi, S.J.***, Viana, J., Dempster, E.L., Crawford, B., Rutter, M., Mill, J. and Sonuga-Barke, E.J., 2016. Severe psychosocial deprivation in early childhood is associated with increased DNA methylation across a region spanning the transcription start site of CYP2E1. *Translational psychiatry*, *6*(6), p.e830.

Chapter 3

Marzi, S.J., Sugden, K., Arsenault, L., Belsky, D.W., Burrage, J., Corcoran, D., Danese, A., Fisher, H., Hannon, E., Moffitt, T.E., Odgers, C.L., Pariante, C., Williams, B.S., Wong, C.C.Y., Mill, J., Caspi, A. Analysis of DNA methylation in young people reveals limited evidence for an association between victimization stress and epigenetic variation in blood. *American Journal of Psychiatry* (Under review).

Chapter 4

Marzi, S.J., Meaburn, E.L., Dempster, E.L., Lunnon, K., Paya-Cano, J.L., Smith, R.G., Volta, M., Troakes, C., Schalkwyk, L.C. and Mill, J., 2016. Tissue-specific patterns of allelically-skewed DNA methylation. *Epigenetics*, *11*(1), pp.24-35.

Chapter 5

Marzi, S.J., Ribarska, T., Smith, A.R., Hannon, E., Poschmann, J., Moore, K., Troakes, C., Al-Sarraj, S., Newman, S., Beck, S., Lunnon, K., Schalkwyk, L.C., Mill, J. A genome-wide study of H3K27ac identifies molecular signatures of Alzheimer's disease pathology. (Manusccript in preparation)

List of abbreviations

5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
AD	Alzheimer's disease
ASM	Allele-specific DNA methylation
BA	Brodmann area
Bp	Base pair
CETS	Cell epigenotype specific
ChIP	Chromatin immunoprecipitation
Chr	Chromosome
CpG	Cytosine-phosphate-guanine dinucleotide
DHS	DNase I hypersensitive site
DMP	Differentially methylated position
DMR	Differentially methylated region
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DSM	Diagnostic and Statistical Manual of Mental Disorders
DZ	Dizygotic
eQTL	Expression quantitative trait loci
ERA	English and Romanian Adoption Study
E-Risk	Environmental Risk Longitudinal Twin Study
EWAS	Epigenome-wide association study
GWAS	Genome-wide association study
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HPA	Hypothalamic-pituitary-adrenal
ICR	Imprinting control region
IM	Intermediate methylation
IQ	Intelligence quotient
Kb	Kilobase
LCA	Latent class analysis
LD	Linkage disequilibrium

Mb	Megabase
miRNA	MicroRNA
mQTL	Methylation quantitative trait loci
mRNA	Messenger RNA
MSRE	Methylation-sensitive restriction enzyme
MZ	Monozygotic
ncRNA	Non-coding RNA
Nt	Nucleotides
QC	Quality control
qPCR	Quantitative PCR
PCR	Polymerase chain reaction
RNA	Ribonucleic acid
Rpm	Rotations per minute
RRBS	Reduced representation bisulfite sequencing
SAM	S-adenosylmethionine
SES	Socioeconomic status
SNP	Single nucleotide polymorphism
TET	Ten-eleven translocation
ToM	Theory of Mind
UTR	Untranslated region
WGBS	Whole genome bisulfite sequencing
XCI	X-chromosome inactivation

Table of Contents

Table of Figures

Table of Tables

1. Introduction

This chapter serves to introduce the importance of epigenetics in the context of mental health and psychiatric disease. It provides an overview of epigenetic mechanisms, experimental methodology, epidemiological considerations and presents the main aims of my PhD studies in the broader context of epigenetic epidemiology of mental health across the life-course.

Neuropsychiatric disorders are responsible for an immense burden on affected individuals, their families and the health care system. Globally, mental health and drug abuse disorders are the leading cause of years lived with disability, as shown by a 2013 report (Whiteford et al., 2013). Efforts to identify genetic causes of neuropsychiatric disease have thus far had inconclusive success: several large consortium efforts have led to the identification of robust disease-associated risk variants for diseases including Alzheimer's disease (Lambert et al., 2013), schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014) and depression (Hyde et al., 2016). However, the number of associated loci is limited, effect sizes tend to be small and a large part of the disease heritability remains unexplained. Beyond heritable effects, most complex diseases, including neuropsychiatric disorders, exhibit discordance between monozygotic twins and contributions to disease-variation not explained by genetic factors (**Figure 1-1**), indicating that factors beyond genetic variation contribute to phenotypic variation. As will be discussed, epigenetics offers a potential mechanism by which environmental influences convey molecular factors contributing to disease risk (Petronis, 2010). There is increasing evidence for epigenetic dysregulation associated with neuropsychiatric disease (Lunnon et al., 2014, Hannon et al., 2016a, Sun et al., 2016) and psychosocial environmental factors that are risk factors for disease (McGowan et al., 2009, Houtepen et al., 2016, Kumsta et al., 2016). Identifying a mechanistic role of epigenetic factors in psychiatric disease will be invaluable for understanding disease aetiology and discovering protective factors or novel psychopharmacological treatments (Boks et al., 2012).

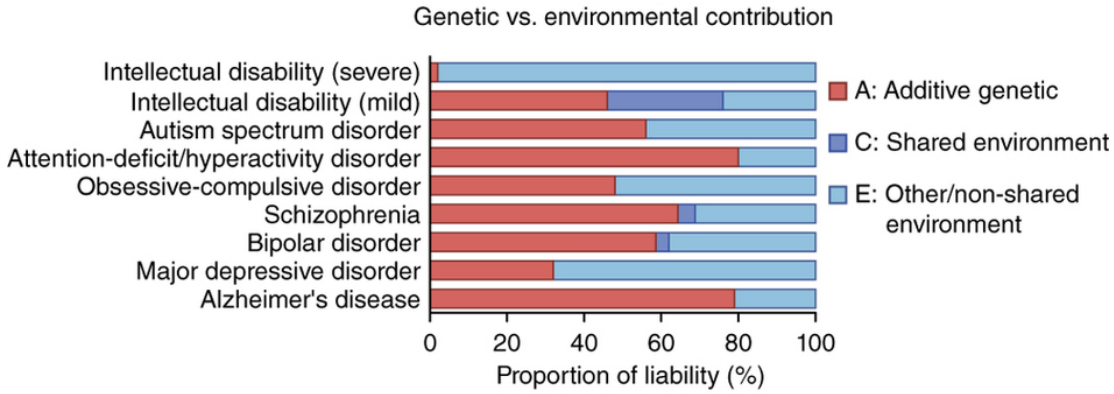


Figure 1-1. Genetic and environmental contributions to neuropsychiatric diseases. Shown are the additive genetic (A), shared environmental (C) and non-shared environmental (E) contribution estimates from large-scale twin or population-based studies (Gandal et al., 2016).

1.1 Epigenetic mechanisms

With few exceptions, all cells in an organism contain the exact same genetic sequence, created at meiosis and remaining stable throughout the life-course. Yet, cells differ substantially in their phenotypic properties and the functions they fulfil in the living organism. The additional information required to generate the adult human body across development, from a single fertilised egg to a fully differentiated organism, is encoded not only genetically but also epigenetically (Waddington, 2014, Wu and Sun, 2006).

The term ‘epigenetics’ was first used in 1957 by Conrad Waddington, who used it to refer to the establishment of cell-fates throughout development, leading from and undifferentiated mass to a complex organism with different organs and tissues (Waddington, 2014). To further illustrate this concept, Waddington used the metaphor of a ball rolling down a landscape of valleys and peaks, each turn determining its cellular-fate one step further (**Figure 1-2**). A more current and functional definition was formulated by the NIH Roadmap Epigenomics Consortium, describing epigenetics as both heritable changes in gene regulation, or changes in the “transcriptional potential”, which are characterised by long-term stability, even though not necessarily heritable. Of note, heritability, in this context, can refer to both mitotic heritability (i.e. alterations are passed on at cell division) as well as transgenerational heritability (i.e. alterations are passed on to the offspring), although empirical evidence of mammalian transgenerational

epigenetic inheritance is still sparse (Heard and Martienssen, 2014, Dias and Ressler, 2014, Rando, 2016). Biologically, the mechanisms generally encompassed by this definition are chemical modifications to the DNA, including DNA methylation and hydroxymethylation, alterations to chromatin structure via histone modifications as well as non-coding RNAs (Kaikkonen et al., 2011). These three types of epigenetic mechanisms are described in the following subsections and a comprehensive glossary of key epigenetic terms is provided in **Table 1-1.**

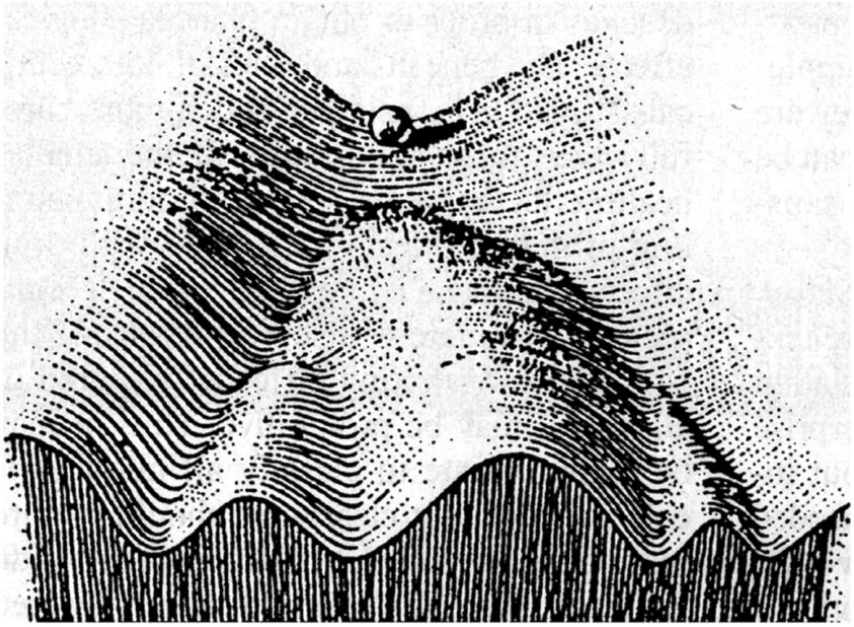


Figure 1-2. Waddington's Epigenetic Landscape. This image serves as a metaphor for the process of cellular differentiation, in which the marble represents a cell following a developmental trajectory. It passes through valleys and peaks, symbolising different layers of regulation, reaching an increasingly determined and irreversible cell-fate (Waddington, 2014).

Table 1-1. Glossary of key epigenetic terms

Term (Key reference(s))	Definition
Chromatin (Berger, 2007)	The protein-DNA complex into which DNA is packaged in the nucleus. Its structure can be altered via chemical modifications to the DNA or the histone proteins associated with it.
CpG island (Gardiner-Garden and Frommer, 1987)	A region of at least 200bp length with a combined GC percentage more than 50% and an observed/expected CpG dinucleotide ratio greater than 60%.
DNA hydroxymethylation (Kriaucionis and Heintz, 2009, Tahiliani et al., 2009)	A chemical modification of cytosine, occurring primarily in the context of a CpG dinucleotide, in which a hydoxymethyl group is added at the 5 position of the cytosine pyrimidine ring. Hydroxymethylation is the derivative of oxidised DNA methylation.
DNA methylation (Jaenisch and Bird, 2003)	A chemical modification of cytosine, occurring primarily in the context of a CpG dinucleotide, in which a methyl group is added at the 5 position of the cytosine pyrimidine ring. DNA methylation in certain regions of the genome can facilitate repression of gene expression by displacing transcription factors and attracting methyl-binding proteins.
Epigenetics (Kundaje et al., 2015)	Heritable changes in gene activity and expression (in the progeny of cells or of individuals) and also stable, long-term alterations in the transcriptional potential of a cell that are not necessarily heritable.
Epigenetic inheritance (Richards, 2006)	Epigenetic modifications can be propagated across mitotic division of cells; they are mitotically heritable. Evidence for transmission through meiosis, a phenomenon known as transgenerational epigenetic inheritance is so far inconclusive.
Euchromatin/ active chromatin (International Human Genome Sequencing Consortium, 2004)	Active chromatin is characterised by open, loose binding of DNA and histones allowing the transcriptional machinery to bind and facilitating gene expression.
Genomic imprinting (Davies et al., 2005)	Monoallelic expression of a subset of genes in a parent-of-origin specific manner. The silencing or "imprinting" of one allele is established by allele-specific epigenetic marks set in the germline.
Heterochromatin/ inactive chromatin (Huisinga et al., 2006)	Inactive chromatin is found in a condensed state of DNA tightly wrapped around histone proteins. It is associated with repressed transcription.

Histone modifications (Berger, 2007)	Post-translational modifications made to the N-terminal histone tails, including acetylation, methylation and phosphorylation. Histone modifications are key architects of chromatin structure.
Non-coding RNA (Birney et al., 2007, Kapranov et al., 2007)	RNA molecules which are not translated into proteins but tend have structural or regulatory roles.
Nucleosome (Luger et al., 1997)	A single DNA-histone complex. The nucleosome consists of eight histone proteins around which a 147bp section of DNA is wrapped.
X-inactivation (Avner and Heard, 2001)	The epigenetic silencing of one copy of the X-chromosome in females for dosage compensation of sex chromosome genes.

1.1.1 DNA methylation

DNA methylation refers to the addition of a methyl-group ($-CH_3$) to the carbon at position 5 of the cytosine pyrimidine ring (**Figure 1-3**), occurring primarily in the context of cytosine-guanine dinucleotides (CpG) in the human genome (Lister et al., 2009). The resulting modified cytosine base is called 5-methylcytosine. DNA methylation plays a key role in a number of biological and developmental processes, including genomic imprinting (Davies et al., 2005), X-chromosome inactivation (Avner and Heard, 2001), silencing of retroviral transposable elements, maintenance of genomic stability, and long-term regulation of gene expression (Jones, 2012).

In contrast to the DNA sequence, DNA methylation, like all epigenetic marks, is dynamic and characterised by development- and tissue-specific patterns (Horvath, 2013, Davies et al., 2012). Establishment of DNA methylation is catalysed by a group of enzymes known as DNA methyltransferases (DNMTs) and uses S-adenosylmethionine (SAM) as a methyl donor (Klose and Bird, 2006) (**Figure 1-3**). Two classes of DNMTs have been described: the *de novo* DNA methyltransferases (DNMT3a and DNMT3b) are responsible for establishing methylation marks at previously unmethylated sites (Okano et al., 1999), in cooperation with DNMT3L (Hata et al., 2002). They are of particular importance in creating the developmental trajectories of the human methylome. DNMT1

preferentially targets hemimethylated CpG sites and is thus mainly responsible for maintenance of methylation patterns at cell-division as the main contributor to the mitotic heritability of DNA methylation patterns (Li et al., 1992, Goll and Bestor, 2005, Law and Jacobsen, 2010).

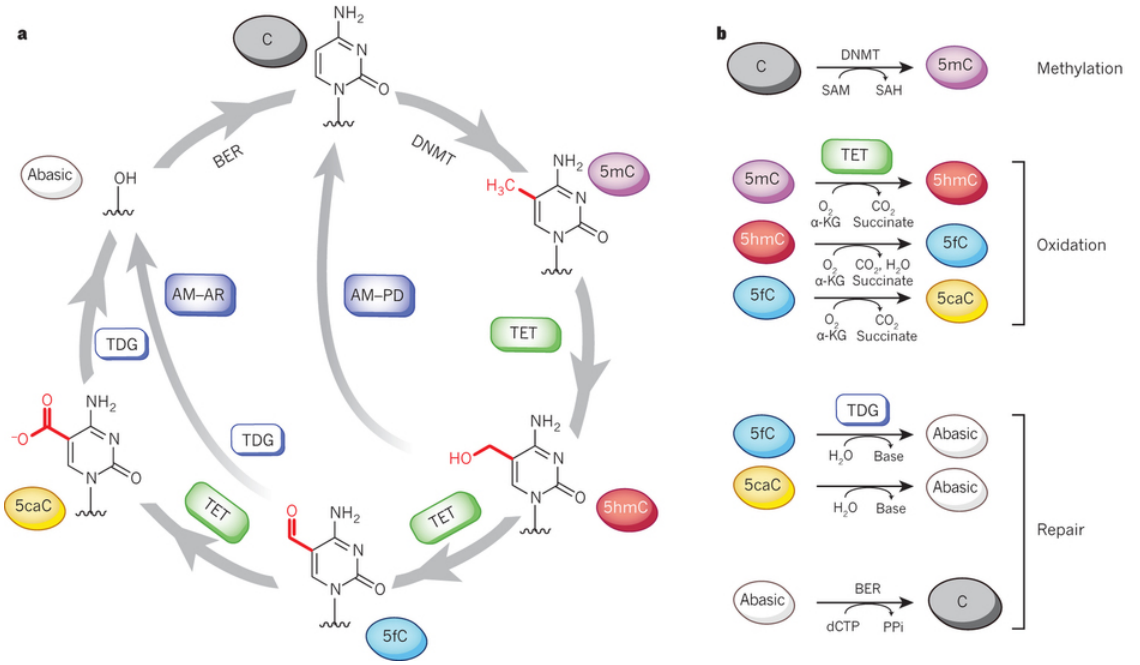


Figure 1-3. DNA methylation and demethylation pathways. a) Overview of the different modified cytosine bases and the enzymes catalysing these reactions. **b)** The addition of a methyl-group at position 5 of the cytosine in a CpG context is catalysed by a group of enzymes called DNA methyltransferases (DNMTs). S-adenosylmethionine (SAM) acts as a methyl donor for this reaction. Oxidation of DNA methylation is catalysed by a group of enzymes known as ten-eleven translocation methylcytosine dioxygenases (TET enzymes). This reaction produces DNA hydroxymethylation, and is the first step in the active demethylation cascade, which also includes the derivatives 5-formylcytosine and 5-carboxycytosine (Kohli and Zhang, 2013).

DNA demethylation takes place via two main pathways: Inactive DNA demethylation describes the absence of maintenance methylation at cell-division, leading to a fully unmethylated strand after the second division. The biochemical processes involved in active DNA methylation are less well characterised. The most prominently described mechanism requires a group of enzymes called ten-eleven translocation methylcytosine dioxygenases (TET enzymes (Wu and Zhang, 2010)). TET enzymes sequentially oxidise methylated cytosine (5mC), resulting in three main methylcytosine derivatives: 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxycytosine (5caC) (**Figure 1-3**).

Deaminated 5hC, 5fC and 5caC can all be removed by thymine DNA glycosylase and are then replaced by unmethylated cytosine as part of the base excision repair pathway (Kohli and Zhang, 2013). While these three intermediates in the 5mC oxidation chain were initially thought to be simply by-products of the demethylation process, more recent evidence suggests functional roles, particularly for 5hmC, which is found at elevated levels in the mammalian brain (Kriaucionis and Heintz, 2009, Globisch et al., 2010, Lunnon et al., 2016).

Although nearly a quarter of methylated cytosines occur in a non-CG context in embryonic stem cells, these disappear almost entirely in differentiated human cells and methylation is limited to the 28 million CpG sites located across the genome (Lister et al., 2009). This number is lower than expected by chance, due to the mutagenic nature of methylated cytosine, which can spontaneously deaminate to thymine (Bird, 1980). Exempt from this global CpG depletion are regions known as “CpG islands”, which are often found in gene promoters and are generally unmethylated (Deaton and Bird, 2011). These regions are characterised by an unusually high density of CpG sites (> 50% CG content, > 60% observed/expected CpG sites) compared to the rest of the genome (Gardiner-Garden and Frommer, 1987). They have been the primary focus of early work in epigenetic studies of complex traits, although increasingly other genomic regions are being highlighted as functionally relevant, including enhancers (Kundaje et al., 2015), CpG shores (Irizarry et al., 2009) and intragenic islands (Illingworth et al., 2010).

Traditionally, DNA methylation has been linked to repression of gene expression, by disrupting transcription factor binding (Watt and Molloy, 1988). However, more recent studies have uncovered a more complex relationship between DNA methylation and gene expression, in which the functional role of DNA methylation depends largely on the genomic context in which it occurs (Jones, 2012). CpG islands occurring at promoters and transcription start sites (TSS) are usually hypomethylated. A few exceptions exist, where promoter CpG island hypermethylation is associated with long term transcriptional silencing. Most notably, this occurs in regions controlling the monoallelic expression of imprinted genes and on inactive X-chromosomes in the context of X-chromosome inactivation (Jones, 2012) and is observed in combination with the repressive H3K9me3 histone modification and methylated DNA-binding proteins, which in

turn recruit histone deacetylases (HDACs), stabilising repression of the region (Illingworth and Bird, 2009, Wade and Wolffe, 2001). DNA methylation at non-CpG island promoters tends to be more variable, and its relationship to gene expression is not yet fully understood. Studies linking DNA methylation at these promoters to repressed expression have thus far been ambiguous (Weber et al., 2005, Gal-Yam et al., 2008). In addition to attracting methylated DNA binding proteins, which initiate transcriptional repression via chromatin compaction (Hendrich and Bird, 1998, Boyes and Bird, 1991), DNA methylation in promoter regions can affect transcription factor binding directly, by disrupting the recognition motifs (Chen et al., 2011, Watt and Molloy, 1988). Beyond promoter methylation, regions up to 2kb upstream of the TSS, known as “CpG island shores” exhibit variable DNA methylation associated with gene expression in a tissue-specific manner (Irizarry et al., 2009).

While gene bodies are generally CpG poor and highly methylated, CpG islands do occur in these regions. Non-promoter CpG islands have been termed “orphan CpG islands” and have been proposed to play a role in development after which they become methylated and lose their promoter function (Illingworth et al., 2010). DNA methylation in gene bodies is positively correlated with gene expression, and suggested functions of methylation in intragenic regions include the silencing of repetitive DNA elements (Yoder et al., 1997), alternative splicing (Shukla et al., 2011) as well as alternative promoter usage (Maunakea et al., 2010). Other regulatory regions at which DNA methylation may play a functional role are enhancers and insulators. Enhancers, short regions of DNA which interact with gene promoters to control gene expression, are CpG poor and are characterised by variable methylation levels (Jones, 2012). While the function of DNA methylation at enhancers is not fully understood, it has been linked to reduced enhancer activity (Schmidl et al., 2009). At insulators - DNA elements which can block interactions between enhancers and promoters – DNA methylation has been linked to instances of differential CTCF-binding. DNA methylation has been shown to inhibit CTCF-binding at specific insulators, and can thereby alter the enhancer-insulator interaction at some loci (Bell and Felsenfeld, 2000); at other loci the resulting absence of CTCF-binding is linked to alternative splicing (Shukla et al., 2011).

1.1.2 Histone modifications

Histone proteins play a key role in DNA packaging and compaction inside the cell nucleus. The compacted compound of DNA and proteins is called chromatin and its basic building blocks are nucleosomes – a complex of approximately 147bp of DNA wrapped twice around a core of eight histone proteins. This positively charged histone octamer core is comprised of two copies of each of four histone proteins: H2A, H2B, H3 and H4 (Spencer and Davie, 1999) (**Figure 1-4**). The "beads on a string" structure resulting from DNA-nucleosome associations is then folded into higher-order chromatin structures (Woodcock and Dimitrov, 2001).

Histone proteins additionally play a role in the regulation of gene expression by altering the physical accessibility of DNA sequences (Jiang and Pugh, 2009, Venkatesh and Workman, 2015). The most active part of chromatin is labelled as euchromatin, representing > 90% of the human genome (International Human Genome Sequencing Consortium, 2004). It is often loosely packed and can facilitate active gene expression by allowing the transcriptional machinery to bind to the DNA. In contrast, heterochromatin is characterised by densely packaged chromatin. It is generally silenced either directly by making DNA inaccessible to the transcriptional machinery or by post-transcriptional silencing mechanisms including RNA interference (Grewal and Jia, 2007, Spencer and Davie, 1999, Volpe et al., 2002) (**Figure 1-5**). The interaction of DNA and histone proteins and the resulting chromatin structure is partly controlled via chemical modifications to the N-terminal tails of the histone proteins, which extend out of the nucleosome complex (Spencer and Davie, 1999). A growing number of histone modifications have been identified, including acetylation, methylation, phosphorylation and ubiquitylation. The aggregate action of these modifications constitutes a complex 'histone code' (Berger, 2007, Ernst and Kellis, 2010, Kasowski et al., 2013). Human tissue-specific variation across a range of histone modifications has recently been profiled comprehensively by the NIH Roadmap Epigenomics Consortium (Kundaje et al., 2015).

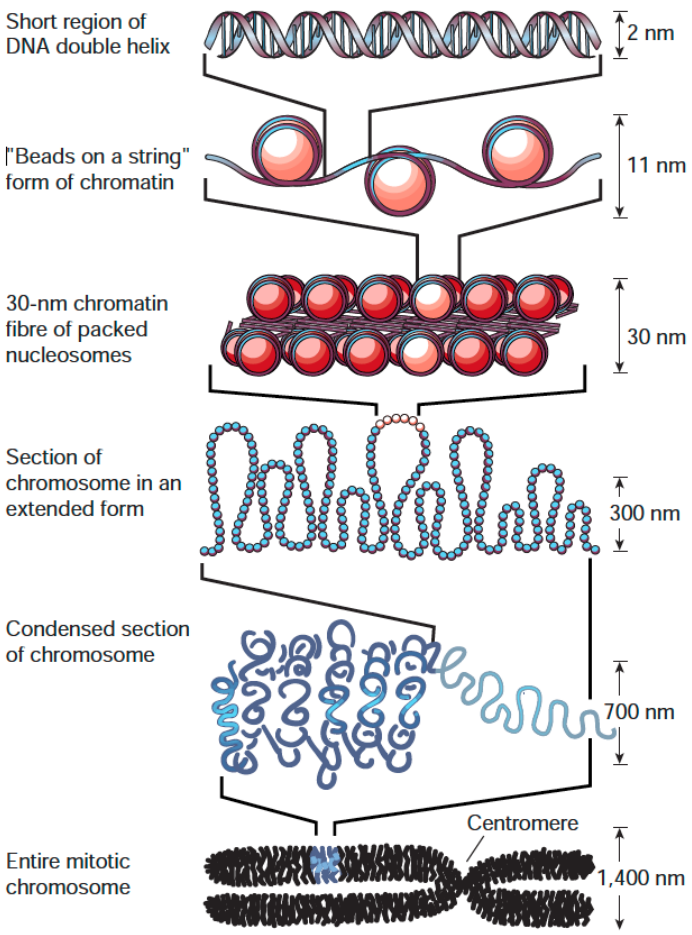


Figure 1-4. Chromatin structure. The nucleosome describes each unit of DNA wrapped around a core composed of histone proteins. It is the smallest and most basic level of chromatin structure and compaction. At the next level this “beads on a string” structure folds to create a 30 nm chromatin fibre, which in turn is folded into higher-order chromatin structures (Felsenfeld and Groudine, 2003).

Lysine acetylation is possibly the best studied histone modification (Strahl and Allis, 2000). It is dynamically modified by the enzymes histone acetyltransferase (HAT) and histone deacetylase (HDAC), adding and removing acetyl groups, respectively (Saha and Pahan, 2006) and is associated with transcriptional activity (Berger, 2007, Strahl and Allis, 2000). Other examples of activating and repressing modifications include H3K4 methylation, which is associated with gene activation, and methylation of H3K9 and 27, which is associated with decreased gene expression (Akbarian and Huang, 2009). The most consistently and coherently occurring combinations of histone modifications have been quantitatively categorised into so-called “chromatin states” (Ernst and Kellis, 2010, Ernst et al., 2011, Ernst and Kellis, 2012), and more recently recalibrated as part of the Epigenomics Roadmap Consortium (Kundaje et al., 2015) (**Figure**

1-6). Each chromatin state shows specific functional annotations and biological roles, including “active TSS”, “weak transcription”, “enhancer”, “heterochromatin” and “quiescent”. Histone 3, lysine 27 acetylation (H3K27ac), a mark found at active promoters and enhancers (Creyghton et al., 2010) could be of particular relevance to gene regulation in health and disease, due to its direct link to active regulatory states and the elevated levels of inter-individual variation observed in H3K27ac at enhancers (Kasowski et al., 2013). Analysis of H3K27ac in post-mortem entorhinal cortex and its association with Alzheimer’s disease is a key focus of my PhD thesis, the results of which are presented in **Chapter 5**.

1.1.3 Non-coding RNA

A third type of epigenetic mechanism, in addition to DNA and histone modifications, is constituted by non-coding RNA (ncRNA). These RNA molecules make up the majority of transcripts (Kapranov et al., 2007) with protein-coding genes only making up a small proportion of the human genome (1-2%). While previously thought to be silent, projects like ENCODE (Birney et al., 2007, Encode Project Consortium, 2012) have highlighted widespread transcription across the human genome, including regions overlapping known protein-coding genes, but for most of which biological functions have not been identified thus far.

NcRNAs can be classified into two groups based on their functional role (Kaikkonen et al., 2011): Infrastructural ncRNAs, which include ribosomal RNA (rRNA), RNAse P RNA, small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), telomerase RNA and transfer RNA (tRNA), are ubiquitously expressed and required for normal cellular functioning (Prasanth and Spector, 2007). Regulatory RNAs are comprised of long non-coding RNAs (lncRNAs, > 200 nucleotides (nt)), microRNAs (miRNAs, 22-23nt), piwi-interacting RNAs (piRNAs, 26-31nt) and small interfering RNAs (siRNAs, 20-24nt) (Kapranov et al., 2007, Kaikkonen et al., 2011).

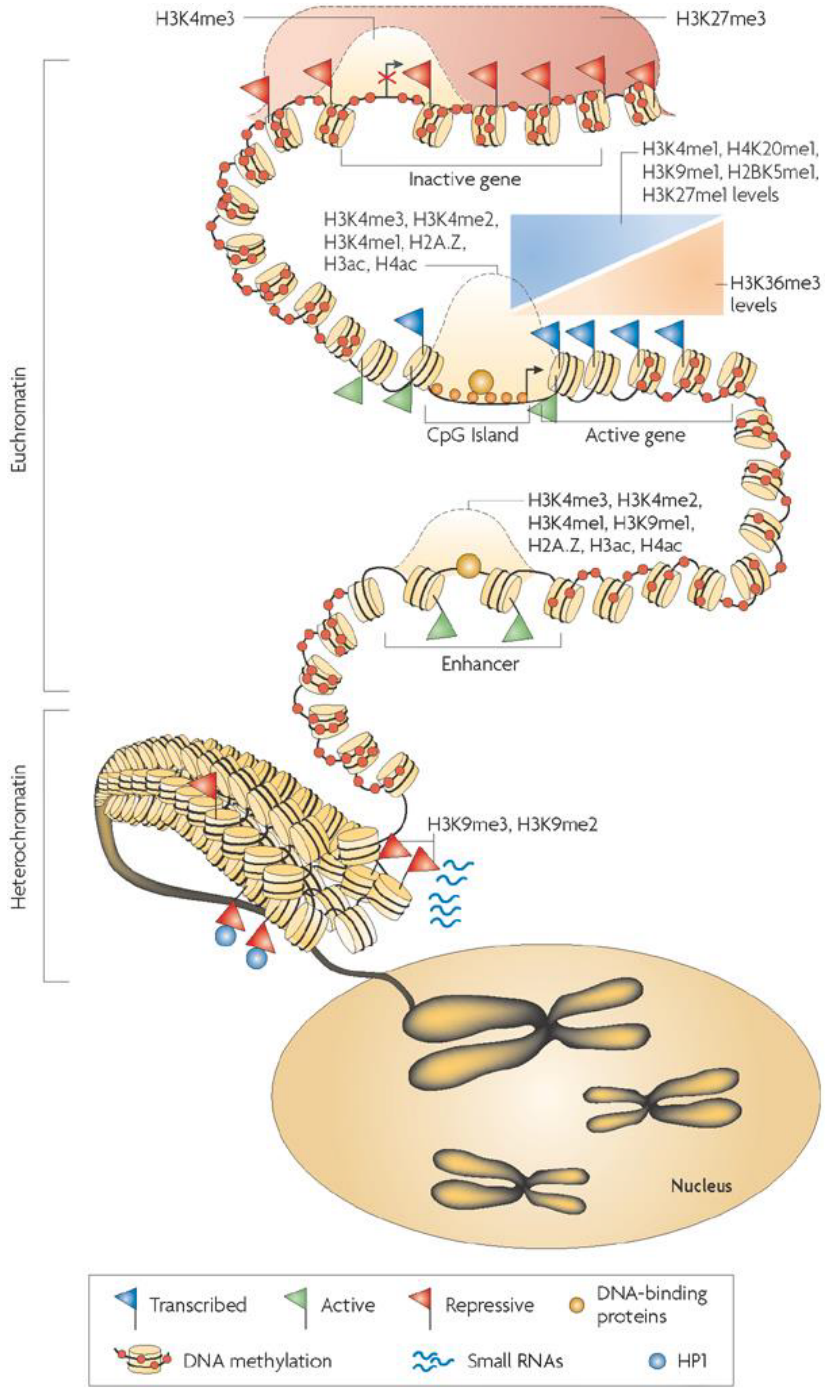


Figure 1-5. Epigenetic modifications interact to steer gene regulation. Epigenetic modifications, including DNA methylation, histone modifications and non-coding RNAs can alter the physical structure of the chromosome. One interaction pathway, for example, is triggered by attracting methyl-binding proteins to CpG sites, mediating the access of transcription factors that drive gene expression. These methyl-binding proteins can in turn recruit further repressive mechanisms including histone deacetylases and chromatin remodelling factors. The two main chromatin states of euchromatin (active) and heterochromatin (inactive) are also illustrated in this figure (Schones and Zhao, 2008).

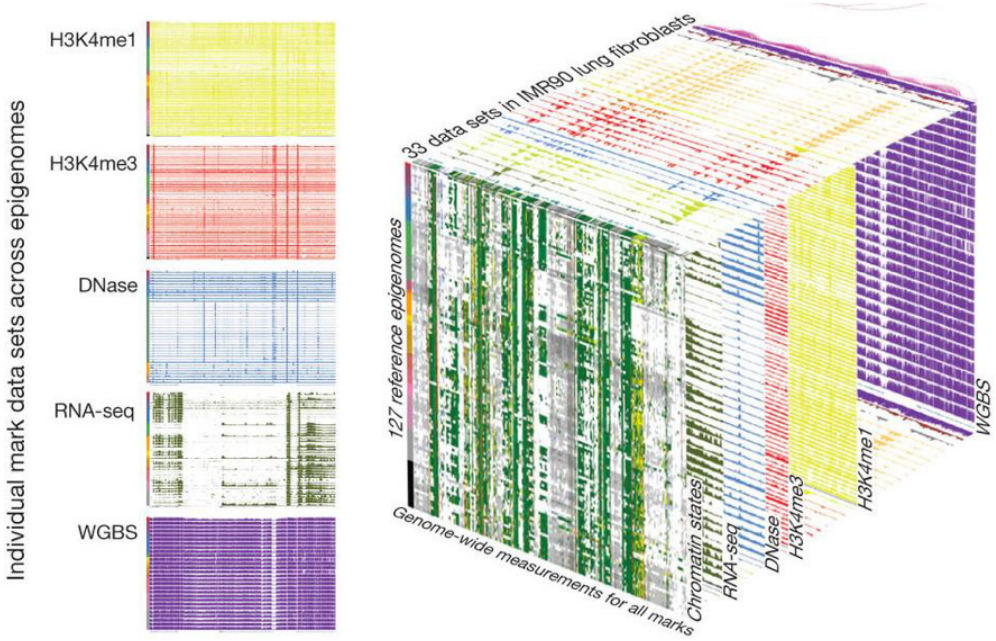


Figure 1-6. Epigenomic information across tissues and marks profiled by the NIH Roadmap Epigenomics Consortium. Various epigenetic modifications were profiled across different tissues and cell types in the human body by the NIH Roadmap Epigenomics Consortium. The left panel illustrates specific individual epigenomic marks across all epigenomes in which they are available; the right panel highlights the three dimensions of the consortium effort: 1) 127 different cell- and tissue-types, 2) profiling across the whole genome, 3) characterising at least five core histone modifications (H3K4me3, H3K4me1, H3K36me3, H3K27me3 and H3K9me3) and other regulatory elements (Kundaje et al., 2015).

Non-coding RNAs have been repeatedly reported to have widespread epigenetic effects (Kaikkonen et al., 2011). miRNAs, for example, are involved in post-transcriptional repression via binding to mRNA of protein coding genes (Bartel, 2009, Krol et al., 2010), while siRNAs use a similar repression mechanism but are additionally linked to the establishment of heterochromatin as further source of transcriptional repression (Kaikkonen et al., 2011). LncRNAs, which represent the majority of ncRNAs, can modify gene transcription via antisense transcription and chromatin remodelling (Mercer et al., 2009, Ponting et al., 2009, Dinger et al., 2008).

1.1.4 Interaction of epigenetic mechanisms

While epigenetic mechanisms are most often studied in isolation, they do not act independently of one another, but in a coordinated fashion, interacting in various ways to regulate gene expression (**Figure 1-5**). Previously mentioned examples

are methyl-binding proteins, which recruit HDACs, modifying histone acetylation (Illingworth and Bird, 2009) or lncRNAs interacting with histone modifications and chromatin remodellers (Dinger et al., 2008). Further examples are observed in unmethylated H3K4 attracting *de novo* methyltransferases to initiate DNA methylation (Ooi et al., 2007), or DNMTs, which in addition to their role in methylating DNA, can recruit HDACs to modify chromatin structure (Jaenisch and Bird, 2003).

The co-occurrence and relationship between epigenetic modifications has been studied in detail as part of the NIH Roadmap Epigenomics Consortium project (Kundaje et al., 2015), which profiled five key histone modifications (H3K4me3, H3K4me1, H3K36me3, H3K27me3 and H3K9me3) in 127 different human cell and tissue samples (**Figure 1-6**). In addition, DNA methylation, DNA accessibility, RNA expression and various histone acetylation marks were profiled in a large subsample, enabling the establishment of high level epigenomic maps and profiling of regulatory elements as well as the quantitative characterisation of coordinated regulatory activity across different epigenetic mechanisms (Ernst and Kellis, 2012).

1.2 Methods for measuring epigenetic variation

1.2.1 DNA methylation

Unlike genotypic information, DNA methylation cannot be quantified by standard molecular biology techniques. DNA methylation is erased by polymerase chain reaction (PCR) as well as cloning in bacteria. Methylation-dependent pre-treatments of DNA are thus generally required to quantify DNA methylation. There are three main classes of pre-treatments: First, sodium-bisulfite conversion is the currently most frequently applied technique for interrogating epigenetic correlates of complex disease, allowing quantitative interrogation of DNA methylation status at single locus level. DNA is exposed to sodium-bisulfite, converting unmethylated cytosine to uracil (and in subsequent PCR to thymine) while 5mC remains unaltered. This technique was used to quantify DNA methylation in **Chapters 2** and **3** of this thesis. Second, genomic DNA can be digested by methylation-sensitive restriction enzymes (MSREs), which target specific DNA motifs containing CpG sites and cut DNA exclusively when the CpG site is unmethylated. Examples of MSREs are *Hpa*II (motif: 5'-CCGG-3'), *Hha*I

(motif: 5'-GCGC-3') and *Aci*I (motif: 5'-CCGC-3'), which were used to digest DNA and identify allelic differences in DNA methylation in **Chapter 4** of the thesis. Third, affinity enrichment uses 5mC-specific antibodies or methyl-binding proteins to identify DNA fragments containing methylated cytosines; most commonly this is carried out by methylated DNA immunoprecipitation (MeDIP).

Following pre-treatment, there exist various experimental approaches to quantify the localisation and occurrence of methylated DNA, which vary with regards to genomic coverage, number of samples, which can be processed, and associated costs. These methods fall into four main categories: locus-specific approaches allow the examination of limited target regions, gel-based approaches rely on gel electrophoresis, array-based approaches quantify methylation at a larger number of sites using pre-defined probe sets on microarrays, and next generation sequencing (NGS) based approaches allow for genome-wide quantification of DNA methylation profiles, using advanced NGS technology. The different combinations of pre-treatment and downstream analysis approaches are summarised in **Table 1-2** (Laird, 2010).

When looking for associations between variation in a phenotype or exposure and variation in DNA methylation, widespread coverage of variably methylated sites in the human genome is of interest. Epigenome-wide association studies (EWAS) use high throughput molecular assays to profile DNA methylation, involving array- and sequencing based technologies. The Infinium technology by Illumina is at present the most commonly used platform for genome-wide methylation profiling, owing to the cost efficiency of profiling large numbers of CpG sites (currently > 850,000) at significantly lower cost than NGS-based approaches. This technology relies on a process involving bisulfite-treatment, followed by microarray-hybridisation. The currently available EPIC array is the third version of this platform and was preceded by the 27K (> 27,000 probes) and 450K (> 450,000 probes) arrays. The Illumina 450K methylation array was used in **Chapters 2** and **3** of this thesis.

While genome-wide technologies offer higher coverage and therefore more information, targeted technologies can be a useful compromise when validating observed results. One example of a location-specific technology is pyrosequencing, in which methylation profiles across CpG sites in a targeted genomic region are measured after bisulfite pre-treatment. This technology was

used to validate DNA methylation profiles for a differentially methylated region (DMR) identified in **Chapter 2**.

Table 1-2. Experimental techniques used to study DNA methylation by pre-treatment approach (adapted from (Laird, 2010))

		Pre-treatment		
		Sodium bisulphite	**Enzyme digestion**	**Affinity enrichment**
Analytical step	**Locus-specific**	• MethyLight • EpiTYPER • Pyrosequencing	*Hpa*II-PCR	• MeDIP-PCR
	Gel-based	• Sanger BS • MSP • MS-SNuPE • COBRA	• Southern blot • RLGS • MS-AP-PCR • AIMS	
	Array-based	• BiMP • GoldenGate • Infinium	• DMH • MCAM • HELP • MethylScope • CHARM • MMASS	• MeDIP • mDIP • mCIP • MIRA
	NGS-based	• RRBS • BC-seq • BSPP • WGSBS	• Methyl-seq • MCA-seq • HELP-seq • MSCC	• MeDIP-seq • MIRA-seq

AIMS: amplification of inter-methylated sites
BC–seq: bisulphite conversion followed by capture and sequencing
BiMP: bisulphite methylation profiling
BS: bisulphite sequencing
BSPP: bisulphite padlock probes
CHARM: comprehensive high-throughput arrays for relative methylation
COBRA: combined bisulphite restriction analysis
DMH: differential methylation hybridization
HELP: *Hpa*II tiny fragment enrichment by ligation-mediated PCR
MCA: methylated CpG island amplification
MCAM: MCA with microarray hybridization
MeDIP, mDIP and mCIP: methylated DNA immunoprecipitation
MIRA: methylated CpG island recovery assay
MMASS: microarray-based methylation assessment of single samples
MS-AP-PCR: methylation-sensitive arbitrarily primed PCR
MSCC: methylation-sensitive cut counting
MSP: methylation-specific PCR
MS-SNuPE: methylation-sensitive single nucleotide primer extension
NGS: next-generation sequencing
RLGS: restriction landmark genome scanning
RRBS: reduced representation bisulphite sequencing
–seq: followed by sequencing
WGSBS: whole-genome shotgun bisulphite sequencing

NGS-based approaches, such as whole-genome bisulfite sequencing (WGBS), allow comprehensive profiling of DNA methylation across the human genome at single-base pair resolution (Lister et al., 2009). However, such methods tend to be inappropriate for epidemiological work, given the large numbers of samples required and the associated costs. Reduced representation bisulfite sequencing (RRBS) (Meissner et al., 2005, Gu et al., 2011) offers a potential compromise: DNA is first treated with restriction enzymes and size selected for small fragments, ensuring that only CpG dense regions are covered by the subsequent bisulfite treatment and sequencing.

1.2.2 DNA hydroxymethylation

The most commonly used interrogation technique for DNA methylation analysis, sodium-bisulfite treatment, cannot distinguish between DNA methylation (5mC) and its oxidised derivate DNA hydroxymethylation (5hmC) (Huang et al., 2010, Jin et al., 2010). Given the abundance of 5hmC in the mammalian brain (Kriaucionis and Heintz, 2009, Globisch et al., 2010), it is desirable for epidemiological profiling of DNA modifications in brain to distinguish 5mC and 5hmC signals, so as not to convolute individual profiles and associations of the two modifications. The recent development of oxidative bisulfite treatment, allows for unique quantification of 5mC by oxidising 5hmC to 5fC prior to bisulfite treatment, and thereby making it susceptible to bisulfite conversion (Booth et al., 2012, Booth et al., 2013). This method has been consolidated with the Illumina 450K methylation array and a measure of DNA hydroxymethylation can be derived by subtracting the pure 5mC signal from the aggregated measure obtained by standard bisulfite treatment (Stewart et al., 2015, Field et al., 2015, Lunnon et al., 2016). This technology was used in **Chapter 5** to quantify pure DNA methylation and hydroxymethylation in post-mortem brain samples as part of the secondary analyses.

1.2.3 Histone modifications

The genomic location of specific histone modifications can also be studied via a range of methods, including targeted, global and genome-wide profiling (Huebert and Bernstein, 2005). While traditional methods involving immunoblotting

(Akbarian and Huang, 2009) or immunohistochemistry (Akbarian et al., 2005) fail to quantify levels of histone modifications associated with specific genomic locations, this can be achieved by chromatin immunoprecipitation (ChIP) techniques. These techniques have certain parallels with the affinity enrichment techniques described for DNA methylation profiling; the main difference being that an antibody is used to target a specific histone modification rather than DNA modification and therefore immunoprecipitation is performed on chromatin, i.e. the DNA – histone complex, rather than purified DNA. In brief, a cross-linking agent, typically formaldehyde, is used to fix DNA-histone interactions. The cross-linked lysate is then fragmented and immunoprecipitated with an antibody specific to the histone modification of interest. The isolated DNA fragments can be quantified by quantitative PCR (ChIP-PCR), microarray-hybridisation (ChIP-chip) or next generation sequencing (ChIP-seq), in analogy to methylation quantification techniques. As with DNA methylation analysis techniques, higher genomic coverage comes at an elevated cost. Traditional ChIP protocols require relatively large numbers of cells (~10^7 cells), which can be problematic if working on limited samples, such as extracted single cell-types or embryonic tissue. Native chromatin immunoprecipitation (N-ChIP) (O'Neill and Turner, 2003, Gilfillan et al., 2012, Brind'Amour et al., 2015), in which the cross-linking step is circumvented, can be performed on substantially lower numbers of cells making it a suitable approach for these scenarios (Buckley, 2007). I optimised and applied traditional ChIP-seq techniques to post-mortem brain samples for the work presented in **Chapter 5**, profiling Alzheimer's disease associated variation in H3K27ac in entorhinal cortex samples.

1.3 Drivers of epigenetic variation

In order to understand the role epigenetic mechanisms play in the aetiology of complex diseases, it is important to characterise naturally occurring intra- and inter-individual variation across the human genome and understand what factors shape epigenetic variation. In this section I will discuss the influence exerted by tissue and cell-types, genetic variation, the environment, stochastic events and development on epigenetic variation.

1.3.1 Tissue-specific epigenetic patterns

Considering the role epigenetic mechanisms play in cellular differentiation, it comes as no surprise that different cell- and tissue-types exhibit strikingly different epigenetic profiles (Kundaje et al., 2015, Encode Project Consortium, 2012). These differences can even be observed between different parts of a complex tissue: In the brain, DNA methylation profiles can distinguish between functionally discrete regions (Ladd-Acosta et al., 2007, Davies et al., 2012) and cell-types (Lister et al., 2013, Kozlenkov et al., 2014, Ziller et al., 2015, Mo et al., 2015). **Chapter 4** highlights the tissue-specificity of allele-specific DNA methylation, providing examples of tissue-specific mQTLs and genomic imprinting. One major question in epigenetic epidemiology, and particularly in the field of psychiatry, is how much we can learn about the affected tissue, in our case the brain, using more readily available surrogate tissue, such as whole blood or buccal swabs. **Chapters 2** and **3** investigate what we can learn about epigenetic signatures of early-life adversity in these two surrogate tissues, while **Chapter 4** contributes to the growing number of maps profiling epigenetic phenomena across multiple tissue-types collected from the same samples (Davies et al., 2012, Hannon et al., 2015, Marzi et al., 2016) to determine intra- and inter-individual variation.

1.3.2 Genetics

While epigenetic processes are defined as occurring independent of the DNA sequence, genetic variation has been shown to have widespread effects on epigenetic mechanisms spanning multiple tissues and different epigenetic modifications (Kasowski et al., 2013, Kilpinen et al., 2013, McVicker et al., 2013, Gutierrez-Arcelus et al., 2013, Hannon et al., 2016b). For DNA methylation, the most direct genetic effect is observed in polymorphic CpG sites. If either of the two bases mutates, methylation does not occur. Given the increased mutational potential of methylated cytosines (Bird, 1980), this substitution can even occur somatically during development. Pervasive effects on epigenetic modifications can be observed when genetic variation occurs in and disrupts epigenetic regulators. These disruptions tend to have genome-wide effects with generally severe developmental consequences: mouse knock-down models of methyltransferases result in embryonic death (Li et al., 1992), while mutations of

the X-linked methyl CpG binding protein 2 gene (*MECP2*), cause Rett Syndrome, a progressive neurological disorder, resulting in severe mental retardation as well as physical disability (Amir et al., 1999).

Finally, epigenetic variation can be influenced by genetic variants in *cis* or *trans*. Single nucleotide polymorphisms (SNPs) associated with methylomic variation are termed methylation quantitative trait loci (mQTLs/ meQTLs) in analogy to expression quantitative trait loci (eQTLs). These DNA variants are widespread, often overlap with eQTLs and are the main contributors to allelic imbalances in DNA methylation (Kerkel et al., 2008, Schalkwyk et al., 2010, Tycko, 2010, Shoemaker et al., 2010, Gertz et al., 2011, Bell et al., 2011, Hannon et al., 2016b). Genetic effects on the epigenome provide a possible link between non-coding genetic variation and phenotypic effects, in which epigenetic alterations act as endophenotypes (Meaburn et al., 2010, Wagner et al., 2014, Gutierrez-Arcelus et al., 2013). In fact, it has been suggested that rather than disrupting the sequence of protein-coding sequences, most common variants associated with complex phenotypes have effects on gene regulation (Nicolae et al., 2010, Maurano et al., 2012). A systematic investigation of allelic-skewing across human brain regions and blood is reported in **Chapter 4**, highlighting the prevalence and tissue-specificity of genetically driven DNA methylation patterns (Marzi et al., 2016).

1.3.3 Environmental influences on the epigenome

Despite notable genetic effects on epigenetic variation, various other factors contribute to epigenetic variation in humans. Several studies profiling monozygotic (MZ) twins have shown widespread epigenetic differences between genetically identical twins (Fraga et al., 2005, Kaminsky et al., 2009, Wong et al., 2010), supporting a role of environmental influences on epigenetic variation.

There is growing evidence that specific environmental exposures can have effects on the human epigenome, and the effects of environmental toxins are now being widely studied, including tobacco smoke (Zeilinger et al., 2013, Elliott et al., 2014, Joehanes et al., 2016, Tsaprouni et al., 2014), bisphenol A (Dolinoy et al., 2007, Singh and Li, 2012, Mileva et al., 2014) and air pollution (Baccarelli et al., 2009, Yauk et al., 2008). A recent large-scale meta-analysis of DNA methylation signatures in cigarette smoking spanning 16 cohorts and a total of 15,907 DNA

samples found 2,623 CpG sites associated with cigarette smoking (Joehanes et al., 2016), making it one of the most robust methylation-associated environmental stimuli to be identified to date. In the context of epigenetic epidemiology, this has the potential to be an important confounder: Patients diagnosed with schizophrenia, for example, are more likely to smoke, which can confound the disease-associated epigenetic signal, if not adequately accounted for (Hannon et al., 2016a). Similarly, other environmental factors can be associated with smoking, including socioeconomic status and early-life victimisation exposure. **Chapter 3** provides evidence that the association between victimisation exposure in adolescence and differential DNA methylation is confounded by cigarette smoking.

In addition to biological environments, psychosocial exposures have been reported to be associated with epigenetic variation. Early-life adversity has been shown to be associated with DNA methylation changes experimentally in rodent models (Weaver et al., 2004, Roth et al., 2009, Kember et al., 2012) as well as observationally in human studies (McGowan et al., 2009, Labonte et al., 2012, Suderman et al., 2012) (see **Chapter 3** for a more detailed discussion of the literature). Given the mitotic heritability of epigenetic modifications, the epigenome is a promising candidate biological mediation mechanism linking these early-life exposures with the observed persistent mental and physical health sequelae. **Chapters 2** and **3** target the question of epigenetic variation associated with early-life adversity using two complementary designs: a natural experiment and a population-based longitudinal cohort study.

1.3.4 Stochastic events

Stochastic variation can contribute to intra- and inter-individually observed variation in the epigenome (Feinberg and Irizarry, 2010). X-chromosome inactivation, the epigenetically regulated silencing of one of the two female X-chromosomes to compensate the dosage of X-linked genes, is the most well-documented form of random epigenetic regulation (Lee and Bartolomei, 2013, Cotton et al., 2015). X-inactivation is a further cause of allelic differences in DNA methylation profiles in females and X-linked genes were identified as allele-specifically methylated by the work presented in **Chapter 4**. Random monoallelic gene expression, resembling X-chromosome inactivation, has been also

observed in autosomes (Jeffries et al., 2012, Gendrel et al., 2014, Eckersley-Maslin et al., 2014, Adegbola et al., 2015), invoking the question of whether random allelic epigenetic processes are widespread in the genome.

1.3.5 Ageing and development

Considerable epigenetic variation can be observed in association with developmental features, most notably chronological ageing and embryonic development. Several studies have identified age-related global hypomethylation across several tissues (Drinkwater et al., 1989, Fuke et al., 2004, Heyn et al., 2012). Site-specific methylomic variation associated with chronological age has also been characterised in a number of studies (Gentilini et al., 2013, Heyn et al., 2012) and shows particularly striking alterations across embryonic development (Spiers et al., 2015). The link between DNA methylation and chronological age is so robust, that researchers have developed a 'DNA methylation clock', a highly accurate tissue-wide epigenetic predictor of chronological age in humans, based on methylation profiles across ~300 CpG sites (Horvath, 2013).

1.4 Epidemiology

1.4.1 Epigenetics of complex diseases

Consistent functioning of mechanisms establishing and maintaining epigenetic profiles is essential for the healthy development and functioning of a differentiated organism (Robertson, 2005). Hence, when key epigenetic regulators are affected, the downstream epigenetic consequences tend to be severe and include neurodevelopmental disorders like Rett Syndrome, caused by mutations in *MeCP2* (Amir et al., 1999), as well as immunodeficiency, centromeric region instability and facial anomalies syndrome, driven by mutations in DNMT3b (Brown et al., 1995, Xu et al., 1999). Imprinting disorders are a second category of severe neurodevelopmental disorders involving epigenetic mechanisms. The best-described case is the Prader-Willi/Angelman syndrome, where a chromosomal deletion in an imprinted gene cluster on chromosome 15 leads to two distinct disease phenotypes, depending on whether it is paternally or maternally inherited (Buiting et al., 1995, Cassidy and Schwartz, 1998). These parent-specific genetic effects arise because imprinted genes are expressed in a

parent-of-origin specific manner, which is controlled by epigenetic silencing mechanisms (Ferguson-Smith, 2011).

Extensive alterations in DNA methylation are observed in cancer, and it is now assumed that both genetic and epigenetic factors contribute to the onset and progression of cancer. Typically, cancer-associated epigenetic variation leads to the silencing of tumour suppressor genes while at the same time upregulating oncogenes, which promote tumour-growth and unconstrained cell proliferation (Sharma et al., 2010).

In recent years, there has been widespread interest in the role epigenetic factors play in the aetiology of common and complex diseases (Rakyan et al., 2011, Petronis, 2010). There is increasing evidence for epigenetic dysregulation in a range of neuropsychiatric phenotypes including schizophrenia (Dempster et al., 2011, Aberg et al., 2014, Pidsley et al., 2014, Hannon et al., 2016a), autism spectrum disorder (Zhu et al., 2014, Sun et al., 2016), depression (Uddin et al., 2011, Murphy et al., 2017, Davies et al., 2014), bipolar disorder (Dempster et al., 2011, Xiao et al., 2014) and attention deficit hyperactivity disorder (Xu et al., 2015, Wilmot et al., 2016). Perhaps the best evidence to date has been found for Alzheimer's disease, with cortex-specific methylomic deregulation patterns identified robustly across multiple independent cohorts (Lunnon et al., 2014, De Jager et al., 2014). Interestingly, most of these diseases share characteristics that are in line with possible epigenetic contributions to the disease aetiology, including MZ twin discordance (**Figure 1-1**), contribution of environmental factors, sex differences in disease prevalence, critical periods of disease susceptibility as well as parent-of-origin specific effects (Mill and Heijmans, 2013, Kaminsky et al., 2006, Kong et al., 2009). Further evidence for a potential mechanistic role of epigenetic dysregulation in complex disease comes from successful trials of drugs altering epigenetic mechanisms (Heerboth et al., 2014). Epigenetic drugs often target broad epigenetic mechanisms and chromatin remodelling agents and the most common types of drugs include methylation inhibitors, HAT inhibitors, and HDAC inhibitors, impacting upon DNA methylation and histone acetylation, respectively. While these drugs are widely used in cancer, increasingly applications involving neuropsychiatric phenotypes are discovered (Szyf, 2015). HDAC inhibitors in particular, are potent candidate drugs for neurodegenerative disease applications and demonstrate widespread effects

on synaptic plasticity and memory formation (Guan et al., 2009, Fischer et al., 2007, Kilgore et al., 2010).

1.4.2 Epigenetic epidemiology in mental health – main challenges

When investigating associations between epigenetic variation and environmental stimuli or phenotypic variation in population-based studies, there are several major challenges. Importantly, epigenetic epidemiology is different from genome-wide association studies (GWAS) in a number of ways, notwithstanding the apparent similarities in technologies and statistical methodology (Heijmans and Mill, 2012, Mill and Heijmans, 2013, Birney et al., 2016). The human genome is stably established in the germline and its sequence will stay constant over a lifetime and across every tissue in the body – apart from rare somatic mutation. This stability comes with key advantages for genetic epidemiology, including lack of confounding by environmental influences, an implicit direction of causality and convenience sampling of DNA from any tissue and at any point in time. Epigenetic modifications can be more accurately characterised as dynamic human phenotypes and therefore study design must consider these potential problems.

Confounding is a major concern and can happen via any influence associated with epigenetic mechanisms. Common confounders are age (Horvath, 2013), sex (Kaminsky et al., 2006), cellular heterogeneity (Houseman et al., 2012, Jaffe and Irizarry, 2014), and substance use, including tobacco (Joehanes et al., 2016) and medication (Menke and Binder, 2014, Boks et al., 2012). If not appropriately controlled for, these factors can lead to seemingly significant results, even if the phenotype or exposure of interest isn't actually associated with epigenetic variation (see **Chapter 1.3** and **Chapter 3**). Batch effects are a common source of spurious associations; they are linked to the experimental procedure (e.g. DNA extraction, bisulfite conversion, microarray batch) and should be accounted for by randomising samples across batches, as well as accounting for experimental batches statistically in the analyses (Leek et al., 2010).

Causal attribution is extremely difficult in epigenetic epidemiology. In addition to potential unobserved confounding factors, there is a potential for reverse causation, where variation in the epigenetic profile is a downstream effect of the disease progression or even disease-related medication. Unfortunately, epigenetic profiles prior to onset of pathogenesis are not always available, but

one useful strategy to improve causal inference is Mendelian randomisation (Relton and Davey Smith, 2012, Davey Smith and Hemani, 2014). Using genetic effects on DNA methylation (mQTLs), genetic variants can be used as instrumental variables, testing whether these variants also predict disease status. Conversely, identifying epigenetic associations which are independent of genetic influence, is a key interest in studies of environmental exposures.

Given the striking tissue-specificity of epigenetic profiles, tissue- and cell-type composition can have substantial effects on epigenetic studies of complex disease and needs to be considered as a potential confounder. When working with heterogeneous tissues like whole blood or brain, composition differences in the cell-types constituting the tissue, will be associated with different epigenetic profiles and can confound an EWAS if they also happen to be associated with the phenotype of interest. Similarly, age-specific epigenetic changes have the potential to confound epigenetic association studies. While chronological age is generally available for epidemiological samples, cell-type proportions are not typically collected. The advent of reference-based estimators of cell-type composition in whole blood (Houseman et al., 2012, Horvath, 2013) and brain (Guintivano et al., 2013), based on DNA methylation profiles, has offered an indispensable tool for epidemiological studies of DNA methylation profiles. These estimators were used in **Chapters 3** and **5** to account for cell-type composition in whole blood and brain, respectively.

Tissues are a twofold challenge in epigenetic epidemiology: Further to confounding due to cellular heterogeneity (Houseman et al., 2012), the tissue-specificity of epigenetic mechanisms implies that disease-specific epigenetic alterations could also be tissue-specific and therefore the disease-affected tissue would be the only informative sample source (Thompson et al., 2010, Lunnon et al., 2014). Regrettably, this is often impossible for neuropsychiatric diseases. Increasingly, brain banks are providing immensely valuable samples for neuropsychiatric research questions (Kretzschmar, 2009); however, they tend to be focused on neurodegeneration and diseases of old age and the number of samples available is often limited in view of the large numbers ideally required for genome-wide epigenomic profiling. For other research questions, including the study of epigenetic consequences of early-life environment, this will remain impossible.

Given all the caveats and challenges to epigenetic epidemiology, it is clear that the 'perfect study' does not exist. To gain an in-depth understanding of epigenetic epidemiology therefore, combinations of different types of studies and replications of results across multiple cohorts are necessary. Moreover, maps of cross-tissue epigenetic covariation in samples collected from the same individuals need to be charted, to calibrate how much we can learn about disease-associated variation in surrogate tissue (Mill and Heijmans, 2013).

1.4.3 Epigenetic Epidemiology – study designs

Experiments are the most controlled form of empirical research. Treatments are randomised across participants and external influences are tightly controlled to cancel out any external effects (Cochran and Cox, 1950). Unfortunately, most research questions from the realm of epigenetic epidemiology cannot be studied in experimental form: Any investigation of human disease will have to be observational (Rosenbaum, 2002). Even amongst environmental stimuli, exposing participants to severely adverse experiences would be highly unethical and studying early-childhood adversity experimentally is unthinkable. An overview of possible study designs to elucidate the role of epigenetic factors in the aetiology of complex diseases across the human life-course is shown in **Figure 1-7** (Mill and Heijmans, 2013).

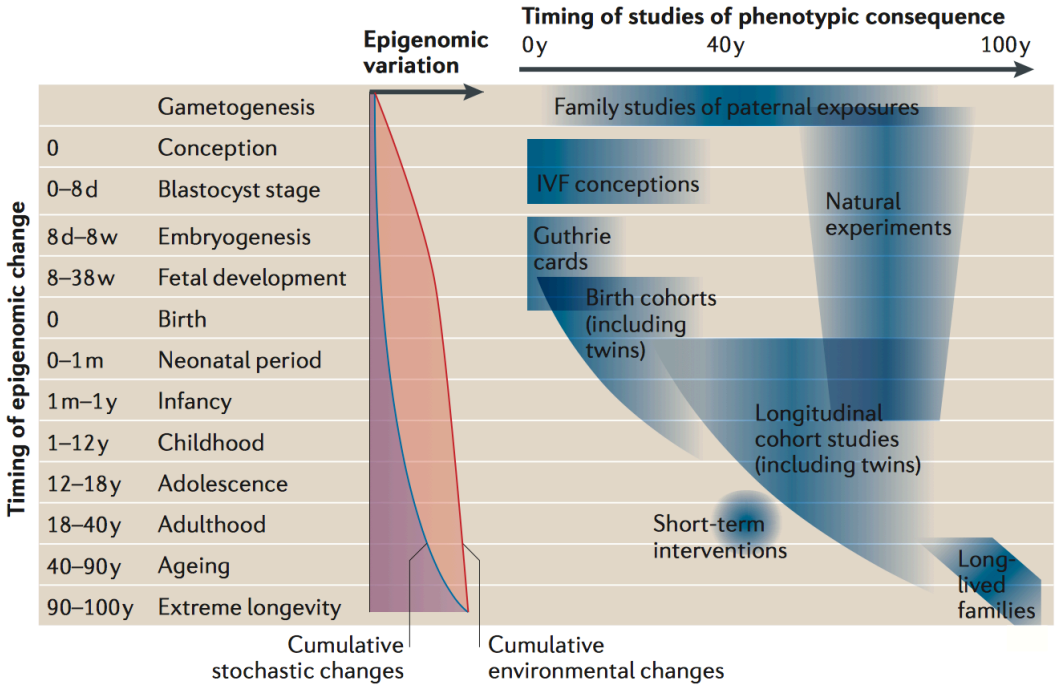


Figure 1-7. Overview of study designs in epigenetic epidemiology. Several study designs help explore the role of epigenetics in complex diseases across the life-course (Mill and Heijmans, 2013).

Experimental epigenetic work is possible in a limited number of scenarios: Animal work has led to some of the most replicated findings in epigenomics of early-life stress using maternal separation and care paradigms (Weaver et al., 2004, Desarnaud et al., 2008, Kember et al., 2012, Kundakovic et al., 2013). In humans, short-term interventions profiling epigenetic marks before and after the intervention can shed light on molecular signatures of therapies or medication (Boks et al., 2012). The closest alternative to a randomised controlled experiment is a *natural experiment*, in which a naturally occurring environmental circumstance is distributed in a quasi-random manner. One of the most prominently studied examples is the Dutch Hunger Winter, following individuals who were in utero during the Dutch famine of 1944 (Lumey et al., 2007). This cohort has provided valuable insights into the sequelae of maternal undernutrition, highlighting stable epigenetic effects (Heijmans et al., 2008, Tobi et al., 2009). The English and Romanian Adoption Study (ERA) is another important natural experiment, following children exposed to severe institutional deprivation in the orphanages of Romania before being adopted into UK families during the fall of the Ceausescu regime in 1989 (Rutter et al., 2010). Given

reported associations between early institutional adversity and intellectual and social behavioural deficits (Kumsta et al., 2015), **Chapter 2** examines associations between deprivation exposure and DNA methylation in adolescence in ERA. An overview of different study designs employed across the four empirical chapters that constitute this PhD thesis is given in **Figure 1-8**.

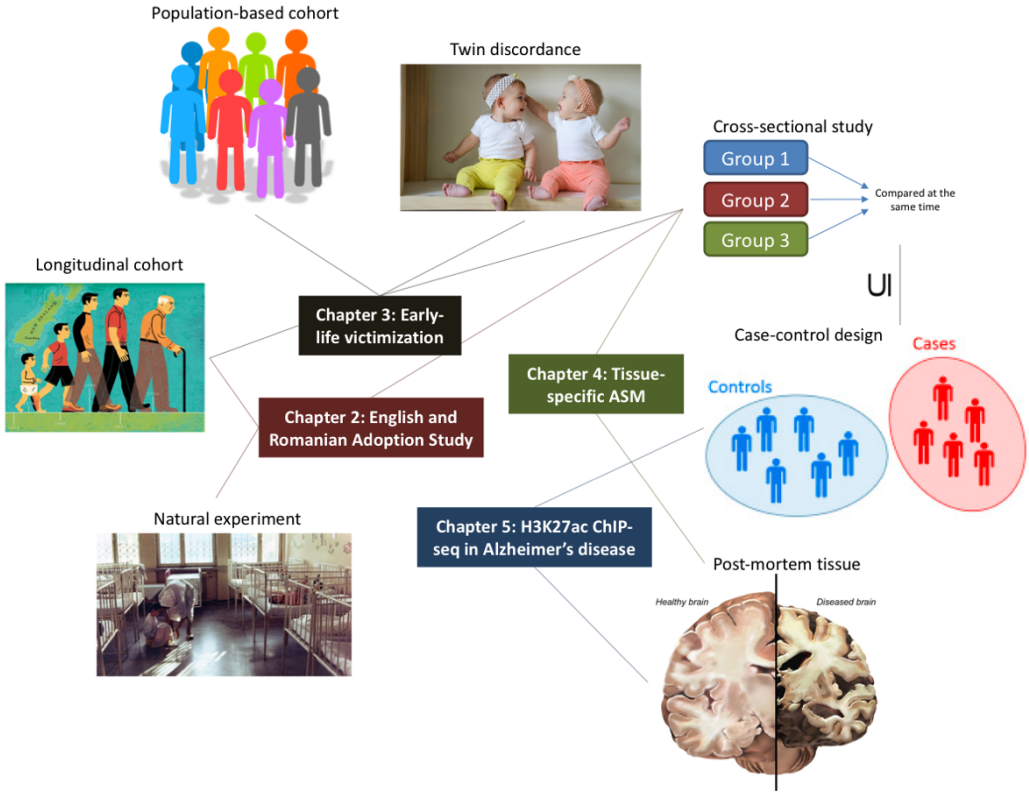


Figure 1-8. Overview of different study designs employed in this thesis. Across the different studies described in this thesis I have employed a range of different study designs to explore the role of epigenetics in mental health from a variety of angles.

Image credits: Natural experiment: Tom Szalay (http://www.npr.org/sections/health-shots/2014/02/20/280237833/orphans-lonely-beginnings-reveal-how-parents-shape-a-childs-brain); Longitudinal cohort: https://moffittcaspi.com/content/aging; Population-based cohort: public domain; Twin discordance: Flickr, Donnie Ray Jones, "Happy Twin Girls"; Case-control design: (Marquand et al., 2016); Post-mortem tissue: Stacy Jannis, Alzheimer's Association (http://discovermagazine.com/2015/march/16-cracking-the-alzheimers-code); all other parts of the Figure were produced by myself.

Longitudinal cohort studies allow the study of epidemiological profiles and epigenetic changes over time. Ideally first sampled during gestation or at birth they allow long-term prospective follow-up of participants' health development, avoiding any sampling biases incurred by retrospective sampling. The Avon Longitudinal Study of Parents and Children (ALSPAC) has investigated

epigenetic trajectories and their relation to health and development of ~14,000 children born in southwest England in the early 1990s (Fraser et al., 2013). Both ERA and the Environmental Risk (E-Risk) Longitudinal Twin Study (Moffitt and E-Risk Study Team, 2002) profiled in **Chapter 3** (**Figure 1-9**) are examples of prospective longitudinal studies of mental health. While ERA focusses on a unique cohort of children exposed to severe early-life adversity in an institutional context, E-Risk consists of a population-based cohort, which is representative of the UK population with new-born infants in the 1990s. Both studies have followed-up participants at regular intervals, with detailed profiling of cognitive, mental and physical health measures.

Despite being examples of longitudinal cohorts, with regard to epigenetic profiling the two aforementioned cohorts represent *cross-sectional studies*; i.e. DNA methylation was profiled at a single time point and is compared across samples rather than longitudinally within the same individuals. This makes causal inference elusive and comes as a drawback of phenotypically deeply profiled longitudinal cohorts, in which biomaterial is only now being collected in an appropriate manner for detailed epigenomic profiling. Following increased interest and widespread research in population health epigenetic epidemiology, these longitudinal studies, including ERA and E-Risk, are starting to collect tissue samples across multiple time points, allowing more precise longitudinal profiling of epigenetic trajectories in the future.

Figure 1-9. A photo collage of twins enrolled in the 'Environmental Risk Longitudinal Twin Study'. Samples from this cohort were profiled for DNA methylation in the study presented in **Chapter 3**.

Studies characterising epigenomic marks across *post-mortem tissue* are inherently cross-sectional as tissue collection is invariably linked to the natural occurrence of death. Therefore, while post-mortem brain research provides extremely valuable insights into tissue-specific epigenetic variation and potentially mechanistic disease-associated epigenetic dysregulation, it will always preclude direct causal attribution, as no temporal sequence of epigenetic variation and phenotypic outcome can be established. **Chapters 4** and **5** characterise epigenomic profiles in post-mortem brain tissue, both in terms of naturally occurring and disease-associated epigenetic variation. A special subgroup of cross-sectional designs are *case-control studies*, in which samples from individuals affected by a disease are compared to healthy control individuals to identify systematic differences (Breslow and Day, 1980). **Chapter 5** represents a case-control design, comparing entorhinal cortex samples from patients affected by Alzheimer's disease (AD) with those obtained from neuropathology-free control individuals for differences in the histone modification H3K27ac.

Given the substantial genetic and environmental influences on the epigenome, *twins* are a powerful approach for controlling confounding factors in epigenetic epidemiology (van Dongen et al., 2012). Monozygotic (MZ) twins share the same genetic sequence, sex, birth date and early-life environments. Thus, epigenetic profiling of MZ twins discordant for diseases or environmental exposures is an effective method to control for variation attributable to these factors. **Chapter 3** uses an embedded twin design, which includes a characterisation of methylomic differences between MZ twins discordant for early-life victimisation exposures (**Figure 1-9**).

1.5 Aims of this thesis

The expanding body of research in epigenetic epidemiology shows promising developments in uncovering molecular biological factors contributing to phenotypic plasticity across a broad range of phenomena. This PhD thesis aims to contribute to this body of research by employing a range of observational epidemiological approaches (**Figure 1-8**) to epigenetic studies of neuropsychiatric phenotypes across the human life-course (**Figure 1-10**). The first two empirical chapters focus on questions of how early-life adversity may

leave stable epigenetic marks across the human body and explore the question whether DNA methylation could act as a biological mediator between the experienced adversity and its associated mental health sequelae. At the same time these two chapters give insights into the utility of two different surrogate tissues (epithelial cells and whole blood). A systematic characterisation of cross-tissue epigenetic variation is undertaken **in Chapter 4**, supplying a new atlas of skewed DNA methylation in the brain and whole blood. Genetic contributions to DNA methylomic variation are investigated in **Chapter 3**, using a discordant MZ twin design, and **Chapter 4**, showing mQTLs as key contributors to allele-specific methylation (ASM). Finally, **Chapter 5** focusses on disease-associated epigenetic variation, profiling AD-affected post-mortem brain tissue in a novel design characterising variation in histone-acetylation.

Aim1:

The first empirical chapter (**Chapter 2**) studies the impact of severe institutional deprivation experienced in the orphanages of the Romanian communist era in the English and Romanian Adoption Study (ERA). Using buccal swab DNA samples collected in adolescence DNA methylation was profiled for 49 individuals using the Illumina Infinium 450K platform. The resulting profiles were examined for differential methylation across probes and regions, associated with psychosocial deprivation exposure and its deprivation-related mental health deficits. This chapter has been published in *Translational Psychiatry* (Kumsta, Marzi et al., 2016).

Aim2:

The association between early-life adversity and DNA methylation is followed up in a large population-based cohort in the second empirical chapter (**Chapter 3**). Using whole blood samples from 1658 individuals, DNA methylation was profiled on the Illumina Infinium 450K platform and associations between multiple victimisation exposures across childhood and adolescence and DNA methylation were examined. In addition to using fine-grained, longitudinally collected phenotypic data, this sample represents one of the largest twin studies of

victimisation exposure, enabling me to control for contributions from genetic effects and familial environments by using an MZ difference design.

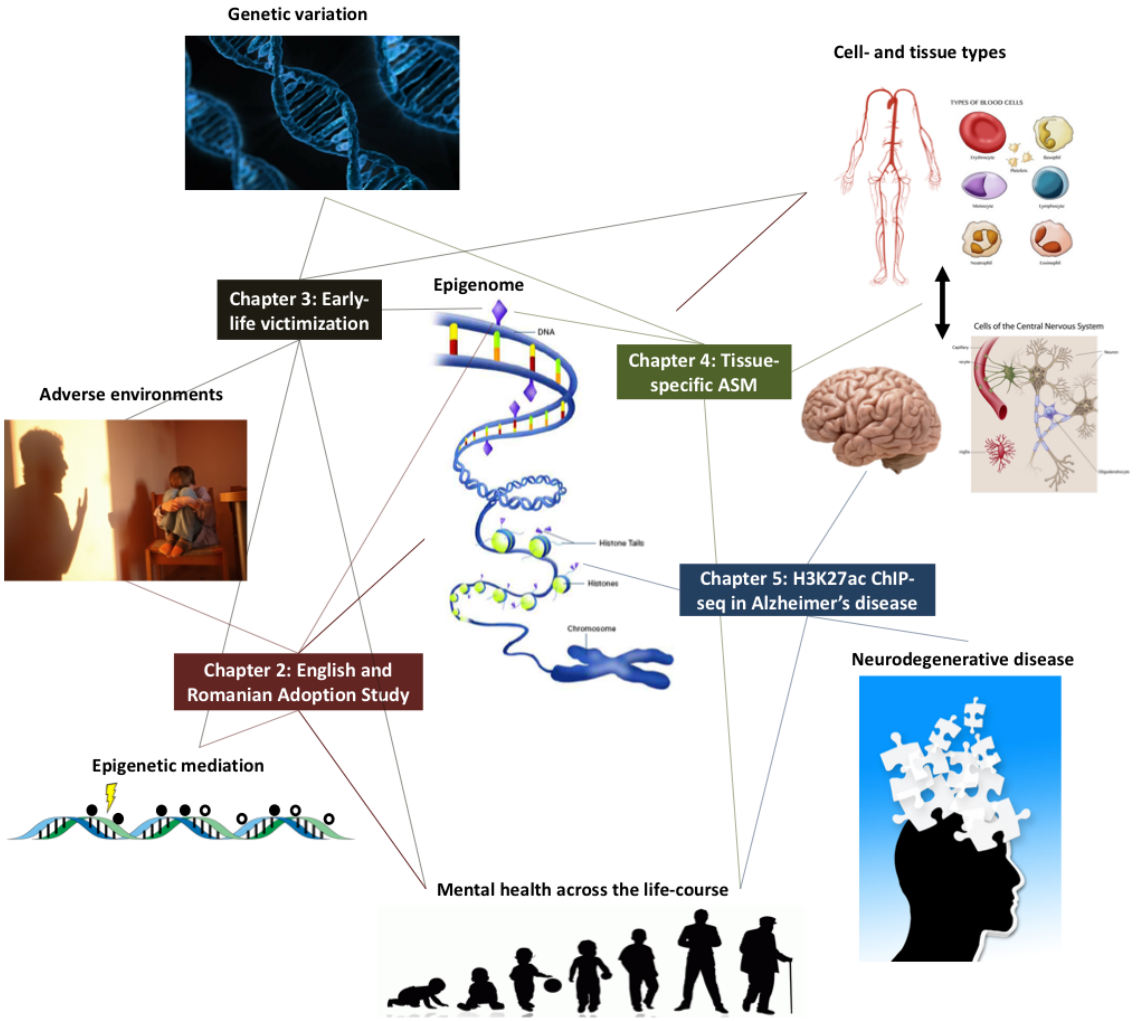


Figure 1-10. Overview of research themes addressed in this thesis. The four empirical chapters presented in this thesis are shown in the context of the research themes they explore. At the centre of the thesis lies the epigenome and its constitutive modifications, and at the basis, all projects involve mental health across the human life-course. While **Chapters 2** and **3** focus on early-life adverse environments and how these can lead to stable adverse mental health outcomes, potentially mediated by epigenetic alterations, **Chapters 4** and **5** focus on mental health in old age and neurodegenerative disease. Other recurring themes are the influence of genetics on the epigenome and the cell- and tissue-type specificity of epigenetic profiles.

Image credits: Epigenome: Crabtree + Company, ID number 2563 (https://images.nigms.nih.gov/Pages/DetailPage.aspx?imageID=313); Adverse environments: Flickr, publik15, "Australia: Kids Help Line says 44 per cent report physical abuse"; Blood cells: Monica Schroeder, "Blood Cell Types"; Brain: Flickr, _DJ_, "human brain on white background"; Brain cells: images.emedicinehealth.com; Mental health across the life-course: http://www.antibullyingpro.com/blog/2015/11/18/what-do-the-latest-stats-show; All other images: public domain.

Aim3:

The third empirical chapter (**Chapter 4**) explores inter- and intra-individual variation in allele-specific DNA methylation (ASM) across the human brain and whole blood. In addition to confirming widespread ASM patterns in blood, this study provides evidence for substantial tissue-specific differences in the amount and location of ASM. This chapter contributes to the mapping of baseline epigenetic variation in the ageing brain as well as the overlap between brain and blood epigenetic patterns. This chapter has been published in *Epigenetics* (Marzi et al., 2016) and tissue-specific ASM patterns are available in a browseable online database: http://epigenetics.essex.ac.uk/ASMBrainBlood.

Aim4:

In the fourth empirical chapter (**Chapter 5**) disease-associated epigenetic variation is examined in post-mortem brain samples from Alzheimer's disease (AD) patients and neuropathology-free control samples. Using chromatin immunoprecipitation followed by sequencing (ChIP-seq) genome-wide profiles of the promoter- and enhancer-associated histone modification H3K27ac were characterised across 47 individuals, providing one of the first genome-wide epidemiological studies of a histone modification to date. In addition to identifying AD-associated variation in a novel epigenetic mark, outcomes were integrated using results from genome-wide association studies (GWAS), known disease-associated proteins, as well as DNA methylation and hydroxymethylation data profiled on the same individuals.

2. Severe psychosocial deprivation in early childhood is associated with increased DNA methylation across a region spanning the transcription start site of *CYP2E1*

2.1 Publication

OPEN

Citation: Transl Psychiatry (2016) **6**, e830; doi:10.1038/tp.2016.95

www.nature.com/tp

ORIGINAL ARTICLE

Severe psychosocial deprivation in early childhood is associated with increased DNA methylation across a region spanning the transcription start site of *CYP2E1*

R Kumsta[1,6], SJ Marzi[2,6], J Viana[3], EL Dempster[3], B Crawford[3], M Rutter[2], J Mill[2,3] and EJS Sonuga-Barke[4,5]

Exposure to adverse rearing environments including institutional deprivation and severe childhood abuse is associated with an increased risk for mental and physical health problems across the lifespan. Although the mechanisms mediating these effects are not known, recent work in rodent models suggests that epigenetic processes may be involved. We studied the impact of severe early-life adversity on epigenetic variation in a sample of adolescents adopted from the severely depriving orphanages of the Romanian communist era in the 1980s. We quantified buccal cell DNA methylation at ~400 000 sites across the genome in Romanian adoptees exposed to either extended (6–43 months; $n=16$) or limited duration (<6 months; $n=17$) of severe early-life deprivation, in addition to a matched sample of UK adoptees ($n=16$) not exposed to severe deprivation. Although no probe-wise differences remained significant after controlling for the number of probes tested, we identified an exposure-associated differentially methylated region (DMR) spanning nine sequential CpG sites in the promoter-regulatory region of the cytochrome P450 2E1 gene (*CYP2E1*) on chromosome 10 (corrected $P=2.98\times10^{-5}$). Elevated DNA methylation across this region was also associated with deprivation-related clinical markers of impaired social cognition. Our data suggest that environmental insults of sufficient biological impact during early development are associated with long-lasting epigenetic changes, potentially reflecting a biological mechanism linking the effects of early-life adversity to cognitive and neurobiological phenotypes.

Translational Psychiatry (2016) **6,** e830; doi:10.1038/tp.2016.95; published online 7 June 2016

INTRODUCTION

Brain circuits underpinning cognition and socioemotional functioning are sculpted by social experiences during early life.[1] Deficient or adverse social environments during this period can increase long-term vulnerability for psychiatric disorders.[2] Understanding the mechanisms linking negative experiences to chronic mental health effects is a key target for translational developmental neurobiology.[3] One hypothesis is that severe social adversity induces long-term alterations to gene expression and function through dynamic epigenetic modifications.[4] Experimental studies in model organisms, for example, have shown that variation in maternal behavior brings about epigenetic alterations and associated changes in gene expression at specific loci, underlying life-long phenotypic differences in physiology and behavior, including neuroendocrine stress responsiveness, fear-related behavior and attentional processes, cognitive development, female reproductive behavior and maternal care itself (see Zhang and Meaney[5] for a review). There is some evidence for similar epigenetic alterations in response to various environmental stressors in humans, including prenatal exposure to famine,[6] psychosocial stress during infancy and pre-school years,[7] early-life socioeconomic status,[8] and childhood abuse.[4,9–11] However, direct replication of the effects observed in experimental animal models in humans remains challenging for a number of reasons. Most sample cohorts are characterised by considerable heterogeneity in the nature, timing and severity of adverse exposures, and there is considerable confounding between early and continuing later adversity and between adversity and consequent mental health problems.

Because of necessary ethical constraints, 'natural experiments'—that is, studies in which exposure to severe adversity is not under direct experimental control—are the best available method for examining epigenetic changes following exposure to severe environmental conditions in human populations.[12] The English and Romanian Adoptees study (ERA) is a prospective longitudinal study of the effects of severe adversity experienced by children before the age of 43 months in grossly depriving Romanian orphanages before they were adopted into UK families at the fall of the Ceauşescu regime in 1989.[13] The children were followed across development and have been assessed at ages 4, 6, 11 and 15 years, with follow-up data collection at the age of 23 years just completed. ERA represents a powerful 'natural experiment' to test the epigenetic hypothesis of the effects of psychosocial adversity. This is because the ERA children (i) were typically exposed to severe deprivation from just after birth for variable, but defined, periods of time (2 weeks to 43 months) and (ii) then experienced a sudden, precisely timed, radical change from a profoundly depriving environment to a nurturing adoptive

[1]Department of Genetic Psychology, Faculty of Psychology, Ruhr-University Bochum, Bochum, Germany; [2]MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; [3]University of Exeter Medical School, University of Exeter, Exeter, UK; [4]Department of Psychology, Institute for Disorders of Impulse and Attention, Developmental Brain-Behavior Laboratory, University of Southampton, Southampton, UK and [5]Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium. Correspondence: Professor EJS Sonuga-Barke, Department of Psychology, Institute for Disorders of Impulse and Attention, Developmental Brain-Behaviour Laboratory, University of Southampton, Building 44, Highfield Campus, Southampton SO17 1BJ, UK.
E-mail: ejb3@soton.ac.uk
[6]These authors contributed equally to this work.
Received 14 January 2016; revised 22 March 2016; accepted 31 March 2016

family one. Furthermore, whereas many in the cohort have displayed long-term persistent deprivation-related problems, at least to adolescence, other adoptees are highly resilient, being indistinguishable in terms of mental health compared with their non-deprived peers. There is a strong association between length of institutional deprivation and risk for persisting deficits. Individuals adopted before 6 months were found to have rates of impairment no different from non-exposed populations, whereas about half of the samples adopted between the ages of 6 and 43 months showed continuing psychological deficits to adolescence.[14] Within the ERA cohort, early adversity is associated with both intellectual and social behavioral deficits, with a characteristic pattern of social impairment across two domains. The first has been termed quasi-autism and is a behavioral pattern characterized by autistic-like features, particularly abnormal preoccupations and intense circumscribed interests. The difference to classical autism lies in greater, albeit unusual, social interest and flexibility, and in the diminishing intensity of these features over time.[15] Deficits in Theory of Mind (ToM) provide substantial mediation of the quasi-autistic pattern.[16,17] The second shares many features with the new DSM diagnostic category 'Disinhibited Social Engagement Disorder'[18] and is characterized by a marked disregard for social boundaries, inappropriate levels of familiarity, social disinhibition and self-disclosure.[15,19] Across all ages there is a substantial overlap between quasi-autism and disinhibited social engagement.[15] These core deficits in social cognition and behavior are often accompanied by cognitive deficits (at the age of 15, the mean intelligence quotient (IQ) of the late adopted group was one s.d. below the UK and early adopted group) and symptoms of attention-deficit hyperactivity disorder.[20]

The persistent nature of the negative impact of early severe deprivation, which for many in the sample was not eradicated by positive post-adoption experiences, is consistent with an enduring biological impact of early deprivation. In this study we aimed to test whether exposure to extreme deprivation is associated with altered DNA methylation among Romanian adoptees and whether these effects are also related to variation in intellectual and social functioning in the ERA group.[17]

MATERIALS AND METHODS

Sample

The ERA sample was drawn from children adopted from Romania into families residing in England between February 1990 and September 1992, who were aged 43 months, or below, at the time of entry to the United Kingdom (see Rutter *et al.*[13] for detailed description of historical background and sample characteristics). Briefly, following an age-stratified sampling design, the ERA study enrolled roughly equal numbers of children adopted before 6 months, between 6 and 24 months and over 24–43 months. The Romanian children were compared with a group of 52 children born and adopted within the United Kingdom before the age of 6 months. None of the children in the within-UK adoptee group had been exposed to early deprivation, neglect or abuse. Most children had been placed in institutions in the first weeks of life (the mean age of entry was 0.34 months, s.d. = 1.26, making it unlikely that the reason for their admission into institutions was manifest handicap). Out of 217 subjects, DNA samples were available for 131 individuals, and 49 individuals were selected for the present study. Our selection strategy was based on the finding that at 11 and 15 years of follow-up there was a step-wise relationship between length of institutional rearing and risk for psychosocial and developmental outcome, with the difference laying between institutional deprivation that did not continue beyond the age of 6 months and institutional deprivation that persisted longer than that.[13] Accordingly, for the current analyses we focused on the comparison between individuals experiencing extended (more than 6 months; $n = 16$) or limited (less than 6 months; $n = 17$) deprivation. Furthermore, these two groups were compared with a subgroup of individuals from the within-UK adoptee group ($n = 16$). Selection of the participants was carried out at random for the UK comparison group and the group experiencing $<$6-month deprivation. For the $>$6-month category, participants were selected at random from the subgroup of individuals showing deprivation-related impairments. As shown in Supplementary Table 1 and Supplementary Table 2, there were no significant differences between groups in gender, smoking or the abuse of alcohol, cannabis or other drugs. With the exception of one individual in the $>$6-month exposure group using antidepressants at time of sampling, and the elevated rate of methylphenidate use in the $>$6-month exposure group, there was no use of medication among the samples included in this study (antipsychotics, mood stabilizers and antidepressants). Furthermore, there were no significant differences in birth weight between the early and late adopted Romanian adoptees. As shown in Supplementary Table 1, and comparable to the ERA sample as a whole,[16] deficits in ToM and lower IQ were observed in the group with extended deprivation. Furthermore, as previously observed,[13] there were no differences between the short length of deprivation and the UK comparison group. The study was approved by the King's College London ethics committee. Parents gave informed consent for themselves and their children.

DNA methylation profiling

Buccal cell samples were collected at the age of 15 and DNA isolated using a standard method.[21] Genomic DNA was treated with sodium bisulfite in duplicate using the EZ-96 DNA methylation-gold kit (Zymo Research, Irvine, CA, USA) and DNA methylation profiled using the Infinium HumanMethylation450 BeadChip (Illumina, San Diego, CA, USA) processed on an Illumina HiScan System (Illumina) according to the manufacturers' standard protocol. All samples were randomized within and between arrays to avoid potential batch effects.

Data-processing and quality control

Signal intensities for each probe were extracted using the Illumina GenomeStudio software and were imported into R (ref. 22) using the *methylumi* [23] and *minfi* package.[24] Multidimensional scaling plots of variable probes on the sex chromosomes were used to check that the predicted gender corresponded with the reported gender for each individual. Stringent quality-control checks, quantile normalization and separate background adjustment of methylated and unmethylated intensities of type I and II probes were implemented using the *wateRmelon* package in R.[25] Samples with $\leqslant$5% of sites with a detection *P*-value $>$ 0.05 were included in subsequent analyses, and probes with $>$5% of samples with a detection *P*-value $>$ 0.05 or a bead count $<$ 3 in 5% of samples were removed. We excluded the 65 single-nucleotide polymorphism probes, probes on sex chromosomes, cross-hybridizing probes[26] and probes with common (minor allele frequency $>$ 5%) single-nucleotide polymorphisms in the CG or single-base extension position from subsequent analysis, with the final analysis data set comprising 382 291 probes.

Cognitive and sociocognitive abilities

Cognitive abilities were assessed with the short form of the Wechsler Intelligence Scale for Children (WISC III, UK) at 15 years of follow-up. This is the most commonly used standardized measure of young people's cognitive abilities, and it has established reliability.[27] Four subscales of the WISC were employed: two from the verbal scales (vocabulary and similarities) and two from the performance scales (block design and object assembly). These four subscales were selected to provide a good estimate of full-scale IQ (reliability coefficient = 0.94).[28] The four subscales were prorated to form a full-scale IQ for subsequent analyses. At the age of 11 years, the 'strange stories' task[29] was employed as a measure of ToM. The task required the children to respond to seven ToM-related vignettes. The responses to the vignettes were scored in terms of the level of ToM understanding displayed, with '0' indicating a non-ToM-related response, '1' indicating basic-level ToM understanding and '2' indicating evidence of more sophisticated ToM understanding. Scores were combined across the seven stories, and the mean scores were used in analyses.

Data analysis

Data was analyzed using a *t*-test for group mean differences in DNA methylation between the two Romanian adoption groups of same ethnicity. No further covariates were included in this test, as all samples were taken at the same age. The potential confounding effect of sex on the identified differences was ruled out through the comparison of results with a sex-regressed model (Supplementary Figure 1). Associations between DNA methylation and exposure time (continuous) as well as IQ and ToM

were analyzed using linear regression. Region-level analysis for deprivation group, ToM and IQ was performed by spatially combining correlated *P*-values using the Python module *comb-p*.[30] We allowed a maximum distance of 1000 bp between neighboring CpG sites, and only included probes with a *P*-value < 0.05 in the initial epigenome wide association scan as starting points for identifying potential differentially methylated regions (DMRs). For each DMR, we report the combined *P*, which is Stouffer–Liptak–Kechris-corrected for regional correlation structure, and the multiple-testing-corrected Šidák *P*-value. The latter corrects the combined *P* for n_a/n_r tests, where n_a is the total number of probes tested in the initial epigenome wide association scan and n_r the number of probes in the given region. The Bioconductor package *bumphunter*[31] was used to confirm DMRs identified by *comb-p* with an alternative method. We report the empirical *P*-value, calculated using 1000 permutations. Genes were assigned to probes using the Genomic Regions Enrichment of Annotations Tool (GREAT) package from the Bejerano Lab at Stanford University (http://bejerano.stanford.edu/great/public/html),[32] taking into account the functional significance of *cis*-regulatory regions.

Bisulfite-pyrosequencing

To technically validate the 450K array data at the *CYP2E1* DMR, a bisulfite-pyrosequencing assay spanning three CpG sites (cg14250048, cg00436603 and cg01465364) was designed using the PyroMark Assay design software (Qiagen, Hilden, Germany). Bisulfite-PCR amplification was performed in duplicate on samples with sufficient remaining DNA using the primers in Supplementary Table 3 and a PCR annealing temperature of 55 °C. DNA methylation was quantified in a subset of 36 samples with sufficient remaining DNA using the Pyromark Q24 system (Qiagen), following the manufacturer's standard instructions, and the Pyro Q24 CpG 2.0.6 software.

RESULTS

Sociocognitive consequences of exposure

Phenotypic analyses on the selected subsample of the ERA cohort used in this study confirmed previously reported negative associations between exposure to severe early-life institutional deprivation and performance in sociocognitive tests (Figure 1). Romanian adoptees exposed to >6 months of deprivation scored significantly lower on tests of both IQ ($P = 0.004$) and ToM ($P = 3.07 \times 10^{-4}$).

Deprivation-associated DNA methylation differences

We first assessed DNA methylation differences at specific 450K array probes between Romanian adoptees categorized as having experienced 'limited' (<6 months in institutional deprivation, $n = 17$) and 'extended' periods of institutional deprivation (>6 months in institutional deprivation, $n = 16$; see Table 1 and Supplementary Figure 2 for the top-ranked differentially methylated positions). No probe-wise differences remained significant after correction for multiple correction, although this is not surprising, given the small number of samples available for this study. DNA methylation differences for the 100 top-ranked exposure-associated differentially methylated positions (Supplementary Table 4) were highly correlated with effect sizes at the same loci in a quantitative analysis of exposure duration ($r = 0.93$, $P = 3.03 \times 10^{-44}$; Supplementary Figure 3), indicating that the effects of severe deprivation at these loci are likely to be cumulative.

We next used *comb-p*[30] to identify spatially correlated regions of differential DNA methylation, identifying a significant DMR on chromosome 10 spanning nine sequential 450K array probes, which were consistently increased in DNA methylation in the severe early institutional deprivation group (combined $P = 2.21 \times 10^{-10}$, corrected Šidák $P = 2.98 \times 10^{-5}$). This region was also identified using an alternative DMR analysis method (*bumphunter*[31]) as showing significantly elevated DNA methylation in the severely exposed group (adjusted $P = 0.002$; Figure 2, Supplementary Figure 4 and Table 2). By comparing the two Romanian adoptee groups with the matched group of children born and adopted within the UK, we were able to show that increased DNA methylation across this DMR is specific to the group that experienced extended deprivation; the control group of UK adoptees was indistinguishable from the Romanian group adopted before the age of 6 months at each of the nine CpG sites comprising the DMR (Figure 2 and Supplementary Figure 4). This ~600bp DMR spans the transcription start site and first exon of the cytochrome P450 gene, *CYP2E1*. There was a significant correlation between DNA methylation levels independently derived from the 450K array and bisulfite-pyrosequencing experiments ($r = 0.52$, $P = 0.001$, Supplementary Figure 5).

Association between DNA methylation and deprivation-related sociocognitive and intellectual impairments

We next tested whether exposure-associated DNA methylation differences were associated with established deprivation-related impairments in cognition and deficits in ToM across samples for

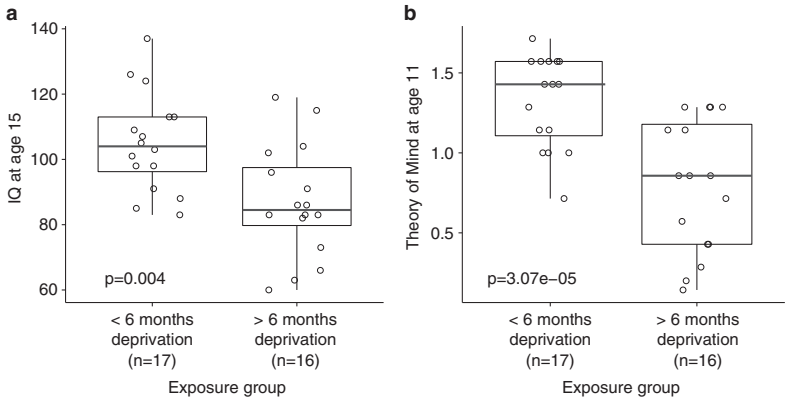


Figure 1. Exposure to severe early-life deprivation is negatively associated with performance in sociocognitive tasks in the English and Romanian Adoptees study (ERA) subsample included in methylomic profiling. Prolonged exposure (⩾6 months) was significantly associated with lower scores for (**a**) intelligence quotient (IQ) at the age of 15 ($P = 0.004$) (**b**) and the Theory of Mind at the age of 11 ($P = 3.07 \times 10^{-4}$).

Table 1. The top-ranked differentially methylated positions associated with exposure to severe institutional deprivation

Probe	*GREAT gene annotation*	P-*value*	*DNA methylation difference*	*>6-Month exposure (mean)*	*<6-Month exposure (mean)*
cg11634248	CHKA, SUV420H1	2.35×10^{-5}	0.03	0.88	0.85
cg14272935	FGF5	6.89×10^{-5}	0.05	0.37	0.32
cg16668903	SNX24, PPIC	9.75×10^{-5}	0.07	0.79	0.72
cg06969206	HHIPL1	1.08×10^{-4}	0.05	0.24	0.19
cg22982014	LGALS4, HNRNPL	1.26×10^{-4}	0.03	0.24	0.21
cg24843511	S100A2, S100A16	1.38×10^{-4}	0.03	0.82	0.79
cg18015809	GPR110, TNFRSF21	1.60×10^{-4}	0.03	0.90	0.88
cg08157194	SLC25A17, MCHR1	1.63×10^{-4}	0.02	0.89	0.87
cg04213775	SLC12A7, NKD2	1.68×10^{-4}	0.03	0.85	0.82
cg07085824	SGK196	1.72×10^{-4}	0.02	0.13	0.10

Abbreviation: GREAT, Genomic Regions Enrichment of Annotations Tool.

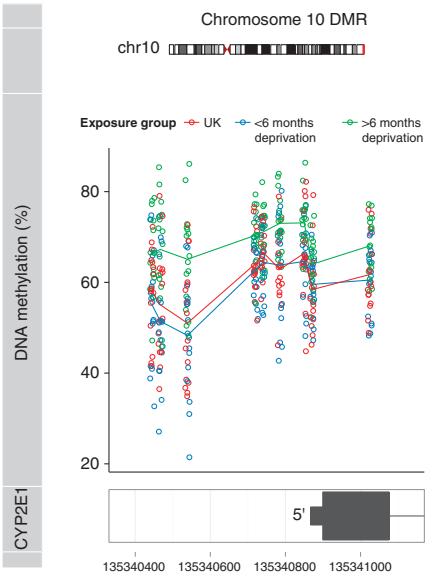


Figure 2. A differentially methylated region (DMR) spanning the transcription start site of *CYP2E1* shows significantly increased DNA methylation levels in adoptees exposed to severe early-life adversity. A DMR on chromosome 10 spanning nine sequential 450K array probes (chr10:135340445-135341026) was identified by *comb-p*. DNA methylation across this region is significantly elevated (combined $P=2.21\times10^{-10}$; corrected Šidák $P=2.98\times10^{-5}$) in individuals exposed to severe long-term (≥6 months) institutional deprivation compared with the low-exposure (<6 months) group and UK control group. Data for an extended region around this DMR are shown in Supplementary Figure 4.

which 450K array data were available. For the 100 top-ranked exposure-group differentially methylated positions, there was a highly significant negative correlation between exposure-group DNA methylation differences and effect sizes at the same probes for both IQ ($r=-0.82$, $P=4.48\times10^{-25}$, Supplementary Figure 6) and ToM ($r=-0.89$, $P=2.23\times10^{-35}$, Figure 3a). Furthermore, using *comb-p* to identify DMRs for sociocognitive and intellectual impairments, we found that DNA methylation across the nine CpG sites in the deprivation-associated *CYP2E1* DMR on chromosome 10 was significantly associated with ToM (combined $P=4.87\times10^{-9}$; Table 3 and Figure 3b) and cognitive impairment (combined $P=2.912\times10^{-5}$).

DISCUSSION

Using samples from a unique 'natural experiment' following children exposed to prolonged severe institutional deprivation, we provide evidence for significant alterations in DNA methylation in response to severe early-life social adversity in humans. We identified a DMR that was associated with extended institutional deprivation across nine adjacent CpG sites spanning the transcription start site and first exon of the cytochrome P450 gene *CYP2E1*. Elevated DNA methylation across this DMR was specific to the group exposed to more than 6 months in Romanian institutions; early-adopted Romanian adoptees were indistinguishable from the control group of UK adoptees—an effect that mirrors prior findings relating deprivation and psychiatric disorders and cognition.[13] DNA methylation across the nine CpG sites in the *CYP2E1* DMR was also associated with ToM performance and cognitive impairment.

The CYP2E1 protein is a member of the cytochrome P450 (CYPs) super family of enzymes, with a role in the metabolism of various exogenous compounds including drugs of abuse and neurotoxins.[33] It is also involved in gluconeogenesis and the synthesis of cholesterol, steroids and other lipids.[34] *CYP2E1* is most abundantly expressed in the liver, although like other CYPs it is present and is active in the human brain, including the frontal cortex, hippocampus, amygdala, hypothalamus and cerebellum (GTEx Analysis Release V4: dbGaP Accession phs000424.v4.p1 (ref. 33)). There is evidence to suggest that CYPs in the brain may have a role in modulating behavior[35] and cognitive processes (for example, shown by imaging genetic studies[36]) as well as susceptibility to central nervous system diseases and drug dependence.[37]

It is currently unknown which molecular pathways in the brain might be affected by changes in *CYP2E1* function, and how deprivation-related sociocognitive deficits might be mechanistically connected to epigenetic variation regulating *CYP2E1*. Of note, increased methylation of a CpG site in close proximity (<1 kb) to our DMR in neonates has been recently associated with prenatal exposure to selective serotonin reuptake inhibitors[38] (Supplementary Figure 4). Although this specific CpG site was not within the DMR identified in our study, it was nominally significantly associated with exposure to adversity ($P=0.034$). Prenatal selective serotonin reuptake inhibitor exposure, similar to severe early-life adversity, has been implicated as a risk factor for long-term cognitive deficits and psychopathology.[39] In a mouse model, it was shown that chronic psychoemotional stress reduced CYP2E1 protein levels by one half, suggesting that stress exposure

Table 2. Association statistics for the nine individual probes in the chromosome 10 DMR

Probe	*Position*	P *>6-month- versus <6-month exposure*	P *UK versus <6-month exposure*	*DNA methylation difference >6-month- versus <6-month exposure*	*>6-Month exposure (mean)*	*<6-Month exposure (mean)*	*UK (mean)*
cg07381788	135 340 445	0.003	0.419	0.12	0.67	0.55	0.58
cg09208540	135 340 467	7.71×10^{-4}	0.408	0.16	0.67	0.51	0.55
cg10986462	135 340 539	3.00×10^{-4}	0.545	0.17	0.65	0.48	0.51
cg01465364	135 340 721	0.001	0.554	0.08	0.70	0.63	0.64
cg00436603	135 340 740	0.007	0.316	0.07	0.71	0.64	0.67
cg14250048	135 340 785	0.010	0.790	0.09	0.73	0.64	0.63
cg19571004	135 340 850	0.003	0.586	0.08	0.73	0.65	0.66
cg19837601	135 340 871	0.053	0.666	0.04	0.64	0.60	0.58
cg21024264	135 341 025	0.004	0.660	0.08	0.68	0.61	0.62

Abbreviation: DMR, differentially methylated region.

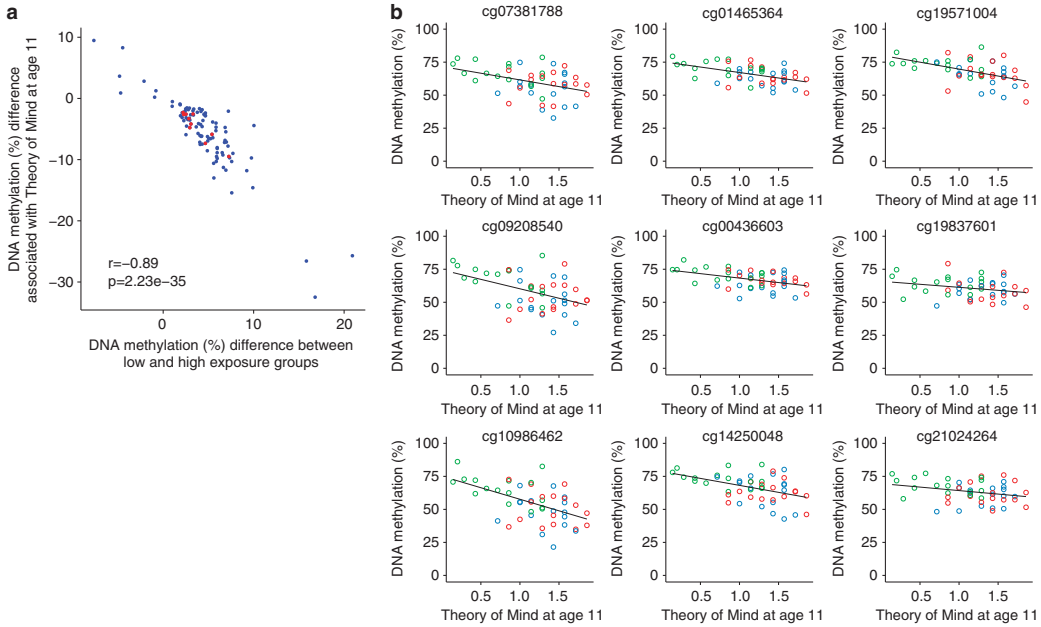


Figure 3. Exposure-associated DNA methylation variation is associated with Theory of Mind test performance. (**a**) For the 100 top-ranked exposure-associated differentially methylated positions (DMPs; see Supplementary Table 4), effect sizes for association with exposure group correlated significantly with effect sizes for association with Theory of Mind at the age of 11 ($r=-0.89$, $P=2.23\times10^{-35}$). The top 10 exposure-associated DMPs (see Table 1) are highlighted in red. (**b**) The *CYP2E1* DMR associated with exposure group is also significantly associated with the Theory of Mind (combined $P=4.87\times10^{-9}$, corrected Šidák $P=8.79\times10^{-5}$). Associations between DNA methylation and Theory of Mind are shown for the nine probes constituting the DMR, with the previously used coloring scheme (UK adoptees: red, short deprivation exposure group: blue, extended exposure group: green).

might have a role in CYP2E1 regulation.[40] Our observation that DNA methylation differences in the regulatory region of *CYP2E1* associated with extended deprivation also correlated with reduced IQ, and ToM is consistent with the view that shared processes may be involved in mediating the observed deficits in both social and intellectual functioning.[17,41]

There is now a large body of evidence and a strong, scientific consensus that childhood stress and early adversity, especially in such extreme forms as the institutional deprivation experienced by the ERA sample, are associated with disturbances of childhood mental health and life-long risks of chronic disorders of mental and physical health.[42] The exact mechanisms by which signals from the social environment impinge on the developing brain to shape the neural circuitry and what role epigenetic processes may have in stabilizing developmental trajectories across the lifespan are only just beginning to be elucidated. Brain structure and function are especially responsive to experience early in life, and development is characterized by the key developmental stages of heightened plasticity.[43] Recent research shows that during these critical periods the genome may be particularly vulnerable to

Table 3. Association statistics of the nine probes in the chromosome 10 DMR with Theory of Mind

Probe	*Position*	P-*value*	*Effect size*[a]
cg07381788	135 340 445	0.006	−0.19
cg09208540	135 340 467	0.001	−0.27
cg10986462	135 340 539	1.10×10^{-4}	−0.32
cg01465364	135 340 721	2.78×10^{-4}	−0.15
cg00436603	135 340 740	0.002	−0.13
cg14250048	135 340 785	0.001	−0.20
cg19571004	135 340 850	3.12×10^{-4}	−0.20
cg19837601	135 340 871	0.072	−0.08
cg21024264	135 341 025	0.052	−0.10

[a]Abbreviations: DMR, differentially methylated region. Effect size is scaled to the maximum Theory of Mind score.

epigenetic disruption.[44] With regard to the ERA sample, it can be speculated that the lack of emotional, sensory and cognitive stimulation associated with deprivation of personalized care during sensitive periods in infant life might have led to epigenetic changes resulting in insufficient fine-tuning of the brain circuitry mediating socioemotional behaviors and underlying higher cognitive function.

One previous study has investigated epigenetic alterations in 8-year-old institutionalized children. Differential methylation patterns (at an extremely non-stringent uncorrected $P < 0.01$) were found at 914 CpG sites, with ~90% of these nominally differentially methylated sites showing elevated methylation in the institutionalized group.[45] In addition to the analytical differences between the studies, samples in this prior study were not exposed to significant deprivation, which may explain the difference in number and magnitude of exposure-associated changes.

Our study has a number of important limitations. First, the number of samples profiled in this study is small, and replication in cohorts with similar types of deprivation experience is warranted. However, the ERA represents a unique natural experiment cohort, and access to equivalent samples exposed to a similar level of adversity for replication is by necessity difficult. Second, the small number of samples means that it is underpowered to formally assess whether deprivation effects on sociocognitive processes are mediated by epigenetic effects. Furthermore, our analyses were cross-sectional, and it is not possible to causally link exposure to the variation we observe. It cannot be ruled out that the observed differences in DNA methylation were caused by deprivation-related impairments observed in the high-risk group, or by other confounding factors such as adverse prenatal conditions (although no differences in birth weight were observed between the long- and short-exposed Romanian adoptees) or medication use. Finally, the observed differences in DNA methylation were observed in buccal cells, and the extent to which peripheral markers index epigenetic variation in central nervous tissue is still debated.[12,46] Of note, buccal cells derive from the same embryonic cells as brain tissue (ectoderm) and have less cellular heterogeneity compared with whole blood.[47] Although there are well-documented tissue-specific differences in DNA methylation, exposure-associated changes in DNA methylation can be identified in many cell types, and peripheral tissues may have some utilities as potential biomarkers of exposure or disease.[12,46,48] Despite these limitations, our data are consistent with the notion that environmental insults of sufficient biological impact during early development might be associated with epigenetic variation detectable in peripheral cells, and provide further support for a role of epigenetic processes in linking the effects of early-life adversity to cognitive and neurobiological phenotypes.

To conclude, children exposed to extreme early institutional deprivation were characterized by significantly increased DNA methylation across a region of the *CYP2E1* gene with putative functional significance for brain function. These findings support the notion that severe social adversity may induce epigenetic variation in human subjects. Future studies should replicate and extend this finding with larger samples and investigate longitudinal changes in DNA methylation over time as a function of post-adversity environments and genomic variations, and relate these to changes in phenotype. It will be important to further investigate the neurobiological significance of these changes by linking DMRs to brain structure and function.

CONFLICT OF INTEREST

EJS Sonuga-Barke obtained speaker fees, consultancy, research funding and conference support from Shire Pharma. He also received Speaker fees from Janssen Cilag and Consultancy from Neurotech solutions, Aarhus University, Copenhagen University and Berhanderling, Skolerne, Copenhagen, KU Leuven. He obtained Book royalties from OUP and Jessica Kingsley. He also received grants from MRC, ESRC, Wellcome Trust, Solent NHS Trust, European Union, Child Health Research Foundation New Zealand, NIHR, Nuffield Foundation, Fonds Wetenschappelijk Onderzoek-Vlaanderen (FWO) and MQ - Transforming Mental Health. The remaining authors declare no conflicts of interest.

ACKNOWLEDGMENTS

The current ERA study is supported by ESRC grant ES/I037970/1 and MRC grant MR/K022474/1 to EJSS-B. RK acknowledges grant support from the Deutsche Forschungsgemeinschaft (DFG, KU 2479/3-1, KU 2479/3-2). SJM is funded by the EU-FP7 Marie Curie ITN EpiTrain (REA grant agreement no. 316758).

REFERENCES

1 Davidson RJ, McEwen BS. Social influences on neuroplasticity: stress and interventions to promote well-being. *Nat Neurosci* 2012; **15**: 689–695.
2 Repetti RL, Taylor SE, Seeman TE. Risky families: family social environments and the mental and physical health of offspring. *Psychol Bull* 2002; **128**: 330–366.
3 Nelson CA 3rd, Bos K, Gunnar MR, Sonuga-Barke EJ. The neurobiological Toll of Early Human Deprivation. *Monogr Soc Res Child Dev* 2011; **76**: 127–146.
4 Turecki G, Meaney M. Effects of the social environment and stress on glucocorticoid receptor gene methylation: a systematic review. *Biol Psychiatry* 2016; **79**: 87–96, in press.
5 Zhang TY, Meaney MJ. Epigenetics and the environmental regulation of the genome and its function. *Annu Rev Psychol* 2010; **61**: 439–466.
6 Heijmans BT, Tobi EW, Stein AD, Putter H, Blauw GJ, Susser ES *et al.* Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci USA* 2008; **105**: 17046–17049.
7 Essex MJ, Boyce WT, Hertzman C, Lam LL, Armstrong JM, Neumann SM *et al.* Epigenetic vestiges of early developmental adversity: childhood stress exposure and DNA methylation in adolescence. *Child Dev* 2013; **84**: 58–75.
8 Borghol N, Suderman M, McArdle W, Racine A, Hallett M, Pembrey M *et al.* Associations with early-life socio-economic position in adult DNA methylation. *Int J Epidemiol* 2012; **41**: 62–74.
9 McGowan PO, Sasaki A, D'Alessio AC, Dymov S, Labonte B, Szyf M *et al.* Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nat Neurosci* 2009; **12**: 342–348.
10 Labonte B, Suderman M, Maussion G, Navaro L, Yerko V, Mahar I *et al.* Genome-wide epigenetic regulation by early-life trauma. *Arch Gen Psychiatry* 2012; **69**: 722–731.
11 Labonte B, Yerko V, Gross J, Mechawar N, Meaney MJ, Szyf M *et al.* Differential glucocorticoid receptor exon 1(B), 1(C), and 1(H) expression and methylation in suicide completers with a history of childhood abuse. *Biol Psychiatry* 2012; **72**: 41–48.
12 Mill J, Heijmans BT. From promises to practical strategies in epigenetic epidemiology. *Nat Rev Genet* 2013; **14**: 585–594.
13 Rutter M, Sonuga-Barke EJ, Beckett C, Castle J, Kreppner J, Kumsta R *et al.* Deprivation-specific psychological patterns: effects of institutional deprivation. *Monogr Soc Res Child Dev* 2010; **75**: 232–247.
14 Kumsta R, Kreppner J, Kennedy M, Knights N, Rutter M, Sonuga-Barke E. Psychological consequences of early global deprivation. *Eur Psychol* 2015; **20**: 138–151.

15 Kreppner J, Kumsta R, Rutter M, Beckett C, Castle J, Stevens S *et al.* Developmental course of deprivation-specific psychological patterns: early manifestations, persistence to age 15, and clinical features. *Monogr Soc Res Child Dev* 2010; **75**: 79–101.
16 Kumsta R, Kreppner J, Rutter M, Beckett C, Castle J, Stevens S *et al.* III. Deprivation-specific psychological patterns. *Monogr Soc Res Child Dev* 2010; **75**: 48–78.
17 Colvert E, Rutter M, Kreppner J, Beckett C, Castle J, Groothues C *et al.* Do Theory of Mind and executive function deficits underlie the adverse outcomes associated with profound early deprivation? Findings from the English and Romanian adoptees study. *J Abnorm Child Psychol* 2008; **36**: 1057–1068.
18 Association AP. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. 5th edn. American Psychiatric Publishing: Arlington, VA, USA, 2013.
19 Zeanah CH, Smyke AT, Koga SF, Carlson E. Attachment in institutionalized and community children in Romania. *Child Dev* 2005; **76**: 1015–1028.
20 Stevens SE, Sonuga-Barke EJS, Kreppner JM, Beckett C, Castle J, Colvert E *et al.* Inattention/overactivity following early severe institutional deprivation: presentation and associations in early adolescence. *J Abnorm Child Psychol* 2008; **36**: 385–398.
21 Freeman B, Smith N, Curtis C, Huckett L, Mill J, Craig IW. DNA from buccal swabs recruited by mail: evaluation of storage effects on long-term stability and suitability for multiplex polymerase chain reaction genotyping. *Behav Genet* 2003; **33**: 67–72.
22 Team RCR. *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria, 2014.
23 Davis S, Du P, Bilke S, Triche TJ, Bootwalla M. Methylumi: Handle Illumina methylation data. *R Package Version 2140,* 2015.
24 Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 2014; **30**: 1363–1369.
25 Pidsley R, Wong CCY, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450 K methylation array data. *BMC Genomics* 2013; **14**: 293.
26 Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW *et al.* Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* 2013; **8**: 203–209.
27 Wechsler D. *Manual for the Wechsler Intelligence Scale for Children*. Psychological Corporation: London, 1992; Vol. 3.
28 Sattler JM. Age effects on Wechsler adult intelligence scale-revised tests. *J Consult Clin Psychol* 1982; **50**: 785–786.
29 Happe FG. An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *J Autism Dev Disord* 1994; **24**: 129–154.
30 Pedersen BS, Schwartz DA, Yang IV, Kechris KJ. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. *Bioinformatics* 2012; **28**: 2986–2988.
31 Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP *et al.* Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int J Epidemiol* 2012; **41**: 200–209.
32 McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 2010; **28**: 495–501.
33 Strobel HW, Thompson CM, Antonovic L. Cytochromes P450 in brain: function and significance. *Curr Drug Metab* 2001; **2**: 199–214.
34 Tanaka E, Terada M, Misawa S. Cytochrome P450 2E1: its clinical and toxicological role. *J Clin Pharm Ther* 2000; **25**: 165–175.
35 Penas-Lledo EM, Dorado P, Pacheco R, Gonzalez I, Llerena A. Relation between CYP2D6 genotype, personality, neurocognition and overall psychopathology in healthy volunteers. *Pharmacogenomics* 2009; **10**: 1111–1120.
36 Stingl JC, Esslinger C, Tost H, Bilek E, Kirsch P, Ohmle B *et al.* Genetic variation in CYP2D6 impacts neural activation during cognitive tasks in humans. *Neuroimage* 2012; **59**: 2818–2823.
37 Ferguson CS, Tyndale RF. Cytochromes P450 in brain: emerging evidence for biological significance. *Trends Pharmacol Sci* 2011; **32**: 708–714.
38 Gurnot C, Martin-Subero I, Mah SM, Weikum W, Goodman SJ, Brain U *et al.* Prenatal antidepressant exposure associated with CYP2E1 DNA methylation change in neonates. *Epigenetics* 2015; **10**: 1–12.
39 Moses-Kolko EL, Bogen D, Perel J, Bregar A, Uhl K, Levin B *et al.* Neonatal signs after late in utero exposure to serotonin reuptake inhibitors: literature review and implications for clinical applications. *JAMA* 2005; **293**: 2372–2383.
40 Maksymchuk O, Chashchyn M. The impact of psychogenic stressors on oxidative stress markers and patterns of CYP2E1 expression in mice liver. *Pathophysiology* 2012; **19**: 215–219.
41 Kumsta R, Rutter M, Stevens S, Sonuga-Barke EJ. IX. Risk, causation, mediation, and moderation. *Monogr Soc Res Child Dev* 2010; **75**: 187–211.
42 Shonkoff JP, Boyce WT, McEwen BS. Neuroscience, molecular biology, and the childhood roots of health disparities: building a new framework for health promotion and disease prevention. *JAMA* 2009; **301**: 2252–2259.
43 Fox SE, Levitt P, Nelson CA. How the timing and quality of early experiences influence the development of brain architecture. *Child Dev* 2010; **81**: 28–40.
44 Fagiolini M, Jensen CL, Champagne FA. Epigenetic influences on brain development and plasticity. *Curr Opin Neurobiol* 2009; **19**: 207–212.
45 Naumova OY, Lee M, Koposov R, Szyf M, Dozier M, Grigorenko EL. Differential patterns of whole-genome DNA methylation in institutionalized children and children raised by their biological parents. *Dev Psychopathol* 2012; **24**: 143–155.
46 Hannon E, Lunnon K, Schalkwyk L, Mill J. Interindividual methylomic variation across blood, cortex, and cerebellum: implications for epigenetic studies of neurological and neuropsychiatric phenotypes. *Epigenetics* 2015; **10**: 1024–1032.
47 Lowe R, Gemma C, Beyan H, Hawa MI, Bazeos A, Leslie RD *et al.* Buccals are likely to be a more informative surrogate tissue than blood for epigenome-wide association studies. *Epigenetics* 2013; **8**: 445–454.
48 Hou L, Zhang X, Wang D, Baccarelli A. Environmental chemical exposures and human epigenetics. *Int J Epidemiol* 2012; **41**: 79–105.

Supplementary Information accompanies the paper on the Translational Psychiatry website (http://www.nature.com/tp)

2.2 Supplementary Material

Supplementary Table 1

Table 2-4 Demographic information on samples. Associations between exposure and cognitive and social cognitive outcome measures are reported.

	>6 months deprivation (n=16)	**<6 months deprivation (n=17)**	**UK (n=16)**	***P* >6 months vs <6 months derivation***
Sex (m/f)	7/9	9/8	10/6	0.732
Mean deprivation time (months)	20.69	3.47	-	1.41×10^{-5}
Mean IQ at age 15	87.00	105.06	105.75	0.004
Mean Theory of Mind at age 11	0.80	1.32	1.38	3.07×10^{-4}

* *P* values were calculated using a Fisher's exact test for sex and a two-sample t-test for deprivation time, IQ and Theory of Mind

Supplementary Table 2

Table 2-5. Sample characteristics for potential confounders – including birth weight and substance use.

	>6 months deprivation (n=16)	**<6 months deprivation (n=17)**	**UK (n=16)**	***P* >6 months vs <6 months derivation***
Birth weight (kg)	2.86	2.96	3.28	0.557
Smoking, alcohol, drug abuse (yes/no)				
Self-report				
Smoking (>5 cigarettes/day)	1/15	2/16	5/16	0.525
Any sign of alcohol abuse	3/16	3/16	4/15	0.641
Occasional cannabis use	1/15	2/15	3/15	0.500
Hard Drug Use	0/15	0/16	2/15	-
Parent report				
Smoking (>5 cigarettes/day)	0/15	0/17	0/16	-
Any sign of alcohol abuse	1/16	3/17	2/16	0.324
Occasional cannabis use	1/16	0/16	0/16	0.500
Drug use	0/15	0/17	0/16	-

* *P* values were calculated using a Fisher's exact test for smoking, alcohol, cannabis and other drug use and a two-sample t-test for birth weight.

Supplementary Table 3

Table 2-6. Primers used for pyrosequencing validation.

Sequencing Primer	Forward Primer	Reverse Primer
TTTTTTAGAATA TATTATAAAATT	5'-GGTATTGGTTG GTGGGTTATT-3'	5'-Biosg-TACTATACACCTACC TCCACATAAACAC-3'

Supplementary Table 4

Table 2-7. Top 100 differentially methylated probes associated with exposure (>6 months vs <6 months exposure). Association between methylation at the top 100 DMPs and deprivation time (quantitative), IQ and Theory of Mind are also reported.

				Exposure group		Deprivation time		IQ		Theory of Mind	
Rank	Probe	Position	GREAT_gene annotation	*P* value	Effect	*P* value	Effect	*P* value	Effect	*P* value	Effect
1	cg11634248	11:67926133	*CHKA, SUV420H1*	2.35×10^{-5}	0.03	0.004	0.04	0.052	-0.06	0.113	-0.03
2	cg14272935	4:81185765	*FGF5*	6.89×10^{-5}	0.05	0.067	0.04	0.250	-0.06	0.021	-0.06
3	cg16668903	5:122206322	*SNX24, PPIC*	9.75×10^{-5}	0.07	0.109	0.05	0.049	-0.13	0.007	-0.09
4	cg06969206	14:100111183	*HHIPL1*	1.08×10^{-4}	0.05	0.005	0.06	0.037	-0.09	0.002	-0.07
5	cg22982014	19:39322497	*LGALS4, HNRNPL*	1.26×10^{-4}	0.03	0.031	0.03	0.020	-0.07	0.002	-0.05
6	cg24843511	1:153579799	*S100A2, S100A16*	1.38×10^{-4}	0.03	0.009	0.03	0.255	-0.03	0.022	-0.03
7	cg18015809	6:47202254	*GPR110, TNFRSF21*	1.60×10^{-4}	0.03	0.070	0.02	0.021	-0.06	0.068	-0.03
8	cg08157194	22:41185264	*SLC25A17, MCHR1*	1.63×10^{-4}	0.02	5.53×10^{-4}	0.03	0.006	-0.06	0.023	-0.03
9	cg04213775	5:1063087	*SLC12A7, NKD2*	1.68×10^{-4}	0.03	0.025	0.03	0.028	-0.07	0.019	-0.04
10	cg07085824	8:42948105	*SGK196*	1.72×10^{-4}	0.02	0.062	0.02	0.136	-0.04	0.102	-0.02
11	cg02132051	14:73020007	*DPF3, RGS6*	1.84×10^{-4}	0.04	0.078	0.03	0.174	-0.06	0.013	-0.06
12	cg24780167	2:179744587	*TTN, CCDC141*	1.98×10^{-4}	0.06	0.041	0.05	0.270	-0.06	0.009	-0.08
13	cg11708358	6:150183516	*LRP11, PCMT1*	2.22×10^{-4}	0.09	0.023	0.09	0.003	-0.25	0.011	-0.12
14	cg14441262	6:33140769	*COL11A2, HLA-DPB1*	2.22×10^{-4}	0.05	0.034	0.04	0.300	-0.05	0.291	-0.03
15	cg08241115	16:722688	*RHBDL1*	2.26×10^{-4}	0.03	0.002	0.04	0.146	-0.05	0.059	-0.03
16	cg21483883	7:37487352	*AOAH, ELMO1*	2.27×10^{-4}	-0.05	0.152	-0.03	0.279	0.05	0.165	0.04
17	cg16699861	16:30429204	*ZNF771, DCTPP1*	2.45×10^{-4}	0.01	0.003	0.01	0.344	-0.01	0.321	-0.01
18	cg21145140	5:110409467	*WDR36, TSLP*	2.69×10^{-4}	0.07	0.004	0.08	0.020	-0.14	0.004	-0.09
19	cg25021532	8:107782513	*ABRA*	2.86×10^{-4}	0.04	0.423	0.02	0.834	-0.01	0.428	-0.02
20	cg14566624	1:6321582	*GPR153*	2.91×10^{-4}	0.07	0.005	0.09	0.078	-0.11	0.032	-0.07
21	cg10986462	10:135340539	*CYP2E1*	3.00×10^{-4}	0.17	0.014	0.19	7.57×10^{-4}	-0.53	1.10×10^{-4}	-0.32

22	cg13628514	12:110271439	*TRPV4*	3.00×10^{-4}	0.06	0.037	0.06	0.810	-0.01	0.105	-0.05
23	cg18167179	14:74058881	*ACOT4*	3.01×10^{-4}	-0.01	0.236	-0.01	0.426	0.01	0.697	0.00
24	cg13443165	9:33130375	*SMU1, B4GALT1*	3.13×10^{-4}	0.21	0.002	0.28	0.062	-0.39	0.022	-0.26
25	cg26385743	2:131101463	*CCDC115, IMP4*	3.18×10^{-4}	0.05	0.088	0.04	0.153	-0.08	0.027	-0.07
26	cg08397920	16:2258580	*C16orf79, MLST8*	3.56×10^{-4}	0.03	0.038	0.03	0.092	-0.05	0.351	-0.02
27	cg07805911	7:818287	*SUN1, PRKAR1B*	3.66×10^{-4}	0.05	0.008	0.07	0.670	-0.02	0.019	-0.06
28	cg22443762	1:197890608	*NEK7, LHX9*	3.82×10^{-4}	0.07	0.053	0.06	0.219	-0.08	0.035	-0.07
29	cg03131097	10:43802881	*FXYD4, RASGEF1A*	4.00×10^{-4}	0.03	0.012	0.03	0.001	-0.09	0.020	-0.04
30	cg20011352	8:37655269	*GPR124*	4.23×10^{-4}	0.02	0.015	0.03	0.148	-0.03	0.001	-0.04
31	cg12748499	1:244374172	*ZNF238, ADSS*	4.32×10^{-4}	0.04	0.013	0.04	0.271	-0.04	0.049	-0.04
32	cg09206774	10:3581233	*PITRM1, KLF6*	4.38×10^{-4}	0.05	0.021	0.05	0.423	-0.04	0.358	-0.02
33	cg16268769	1:156890782	*PEAR1, ARHGEF11*	4.61×10^{-4}	0.04	0.017	0.05	0.041	-0.10	0.002	-0.08
34	cg06705986	10:86004888	*LRIT1, RGR*	4.88×10^{-4}	0.08	0.213	0.05	0.011	-0.18	0.006	-0.10
35	cg18170229	12:322587	*SLC6A12*	5.08×10^{-4}	0.06	0.012	0.07	0.039	-0.11	0.004	-0.09
36	cg13898430	1:25292274	*RUNX3, SYF2*	5.10×10^{-4}	0.04	0.008	0.05	0.016	-0.10	0.050	-0.04
37	cg12973753	6:26208128	*HIST1H4E, HIST1H2BG*	5.17×10^{-4}	0.03	0.019	0.03	0.084	-0.05	0.257	-0.02
38	cg08064292	15:98196234	*ARRDC4*	5.52×10^{-4}	0.10	0.122	0.07	0.130	-0.15	0.067	-0.10
39	cg09000469	10:134229757	*INPP5A, STK32C*	5.53×10^{-4}	0.04	0.035	0.04	0.139	-0.06	0.326	-0.02
40	cg18243598	7:2172011	*MAD1L1, ELFN1*	5.67×10^{-4}	0.02	0.092	0.02	0.049	-0.05	0.180	-0.02
41	cg12637397	5:1229127	*SLC6A18, TERT*	5.72×10^{-4}	0.03	0.158	0.02	0.049	-0.06	0.236	-0.02
42	cg19099736	22:29702810	*GAS2L1*	5.86×10^{-4}	0.03	0.005	0.04	0.398	-0.03	0.338	-0.02
43	cg20203971	2:240171099	*HDAC4, TWIST2*	5.97×10^{-4}	0.02	0.004	0.03	0.005	-0.06	0.009	-0.03
44	cg07335357	17:77758850	*CBX2, CBX8*	6.14×10^{-4}	0.02	0.030	0.02	0.661	0.01	0.341	-0.01
45	cg05914582	12:6588003	*VAMP1, MRPL51*	6.18×10^{-4}	0.03	0.012	0.03	0.003	-0.08	0.002	-0.05
46	cg04522915	20:646942	*SRXN1, SCRT2*	6.59×10^{-4}	0.04	0.036	0.04	0.020	-0.11	0.020	-0.06
47	cg01164118	3:74571289	*CNTN3*	6.87×10^{-4}	-0.05	0.076	-0.04	0.368	0.04	0.730	0.01
48	cg15889768	17:42619247	*GPATCH8, FZD2*	6.92×10^{-4}	0.04	0.076	0.03	0.260	-0.04	0.020	-0.05

49	cg13392468	9:139701199	*C9orf86*	6.95×10^{-4}	0.06	0.162	0.04	0.908	-0.01	0.140	-0.05
50	cg01573067	6:168067266	*TCP10, MLLT4*	7.55×10^{-4}	0.06	0.004	0.08	0.010	-0.14	7.34×10^{-4}	-0.09
51	cg09208540	10:135340467	*CYP2E1*	7.71×10^{-4}	0.16	0.061	0.15	0.004	-0.44	0.001	-0.27
52	cg21648202	19:14171331	*PALM3*	7.89×10^{-4}	0.07	0.115	0.05	0.173	-0.10	0.060	-0.08
53	cg08144503	16:67450471	*ZDHHC1*	7.91×10^{-4}	0.03	0.010	0.04	0.048	-0.07	0.197	-0.03
54	cg23709618	4:119273073	*PRSS12*	8.19×10^{-4}	0.04	0.021	0.05	0.003	-0.12	0.005	-0.06
55	cg07509161	14:54552632	*CDKN3, BMP4*	8.52×10^{-4}	-0.04	0.023	-0.05	0.228	0.05	4.08×10^{-4}	0.08
56	cg00368296	19:8052183	*TIMM44, ELAVL1*	8.66×10^{-4}	0.04	0.132	0.03	0.651	-0.02	0.279	-0.02
57	cg05387404	3:47888039	*DHX30, MAP4*	8.73×10^{-4}	0.03	0.040	0.03	0.395	-0.02	0.038	-0.03
58	cg16005818	1:7700504	*VAMP3, CAMTA1*	8.78×10^{-4}	-0.01	6.12×10^{-4}	-0.01	2.77×10^{-4}	0.03	0.001	0.01
59	cg26380116	12:125927452	*TMEM132B*	8.86×10^{-4}	0.04	0.011	0.05	0.452	-0.04	0.163	-0.04
60	cg17313945	6:32977983	*HLA-DOA*	9.10×10^{-4}	0.05	0.397	0.02	0.274	-0.06	0.037	-0.06
61	cg00476944	7:157411142	*DNAJB6, PTPRN2*	9.35×10^{-4}	0.04	0.038	0.04	0.001	-0.15	0.009	-0.06
62	cg00958409	1:155023062	*ADAM15*	9.39×10^{-4}	0.03	0.010	0.03	0.138	-0.07	0.018	-0.06
63	cg16008138	17:60885892	*MARCH10*	9.88×10^{-4}	0.04	0.083	0.04	0.131	-0.07	0.002	-0.07
64	cg23053407	1:104068312	*RNPC3*	9.89×10^{-4}	0.04	0.042	0.04	0.638	-0.03	0.128	-0.05
65	cg00171161	4:88343636	*NUDT9*	0.001	0.01	0.010	0.01	0.076	-0.02	0.072	-0.01
66	cg04636557	1:231472424	*C1orf124, EXOC8*	0.001	0.07	0.049	0.07	0.404	-0.07	0.009	-0.12
67	cg00053916	8:37457329	*ZNF703, KCNU1*	0.001	0.06	0.016	0.07	0.051	-0.13	0.014	-0.08
68	cg19506311	7:994742	*ADAP1*	0.001	0.07	0.032	0.08	0.022	-0.19	0.032	-0.10
69	cg02334521	10:324041	*ZMYND11, DIP2C*	0.001	0.02	0.015	0.02	0.003	-0.06	0.031	-0.02
70	cg04770705	16:28509428	*APOBR, IL27*	0.001	0.05	0.270	0.03	0.080	-0.10	0.201	-0.04
71	cg03272089	19:2191261	*SF3A2, DOT1L*	0.001	0.02	0.003	0.03	0.001	-0.09	0.194	-0.02
72	cg19786808	17:71431877	*CDC42EP4, SDK2*	0.001	0.04	0.002	0.06	0.033	-0.09	8.51×10^{-4}	-0.07
73	cg18101118	3:45017337	*EXOSC7, ZDHHC3*	0.001	0.07	0.168	0.05	0.069	-0.13	0.161	-0.05
74	cg14375923	13:91827042	*GPC5*	0.001	0.05	0.011	0.06	0.173	-0.08	0.005	-0.09
75	cg06801943	1:2101479	*SKI, PRKCZ*	0.001	0.05	0.047	0.05	0.086	-0.09	0.016	-0.07
76	cg01465364	10:135340721	*CYP2E1*	0.001	0.08	0.007	0.10	0.003	-0.24	2.78×10^{-4}	-0.15

77	cg20610031	4:1950081	*WHSC2, WHSC1*	0.001	0.03	0.002	0.04	0.024	-0.07	0.108	-0.03
78	cg16498741	8:41144520	*ZMAT4, SFRP1*	0.001	0.10	0.039	0.10	0.150	-0.16	0.012	-0.15
79	cg10013969	7:128003601	*PRRT4*	0.001	-0.08	0.026	-0.09	0.021	0.22	0.075	0.09
80	cg03440989	17:40155391	*DNAJC7, CNP*	0.001	0.03	0.066	0.03	0.199	-0.03	0.012	-0.03
81	cg02420027	13:41632640	*WBP4*	0.001	0.04	0.001	0.07	0.438	-0.04	0.472	-0.02
82	cg15426878	3:50368947	*TUSC2*	0.001	0.02	0.017	0.02	0.020	-0.05	0.393	-0.01
83	cg05357287	11:67408159	*TBX10*	0.001	0.07	0.020	0.08	0.843	-0.02	0.246	-0.05
84	cg14868466	10:134018073	*DPYSL4, STK32C*	0.001	-0.02	0.005	-0.03	0.035	0.05	0.026	0.03
85	cg23205676	1:17046897	*NBPF1, MST1P9*	0.001	0.06	0.212	0.04	0.263	-0.08	0.005	-0.10
86	cg11099375	17:62141556	*ICAM2, ERN1*	0.001	0.06	0.006	0.08	0.017	-0.15	7.27×10^{-6}	-0.13
87	cg07617759	2:33050475	*LTBP1, TTC27*	0.001	0.10	0.564	0.03	0.527	0.07	0.431	-0.04
88	cg18524934	2:97164073	*NEURL3, NCAPH*	0.001	0.07	0.078	0.06	0.106	-0.11	0.002	-0.11
89	cg21672450	12:132663793	*NOC4L, GALNT9*	0.001	0.06	0.022	0.07	0.245	-0.07	0.003	-0.10
90	cg26343667	3:156391915	*TIPARP*	0.001	0.03	0.076	0.03	0.278	-0.04	0.333	-0.02
91	cg11516031	4:15776675	*CD38*	0.001	0.02	0.072	0.02	0.015	-0.05	0.004	-0.03
92	cg07813851	1:7466976	*VAMP3, CAMTA1*	0.001	0.05	0.060	0.05	0.680	0.02	0.352	-0.03
93	cg27207776	16:71393801	*CALB2, ZNF23*	0.001	0.06	0.082	0.05	0.069	-0.11	0.233	-0.04
94	cg27028750	5:1349422	*CLPTM1L*	0.001	0.06	0.119	0.04	0.147	-0.09	9.49×10^{-4}	-0.11
95	cg13173415	11:12065669	*MICAL2, DKK3*	0.001	0.05	0.065	0.05	0.111	-0.09	0.018	-0.07
96	cg14186846	6:15256907	*JARID2, DTNBP1*	0.001	0.07	0.091	0.06	0.050	-0.15	0.027	-0.09
97	cg26053073	8:131455383	*ASAP1*	0.001	0.07	0.022	0.08	0.212	-0.12	0.195	-0.07
98	cg24134918	6:149886976	*C6orf72*	0.001	0.05	0.009	0.06	0.017	-0.11	0.296	-0.03
99	cg03511957	1:167765211	*MPZL1, ADCY10*	0.001	0.08	0.013	0.10	0.020	-0.18	0.038	-0.09
100	cg23330788	4:21699988	*KCNIP4*	0.001	0.07	0.225	0.05	0.999	0.00	0.644	-0.02

Supplementary Figure 1

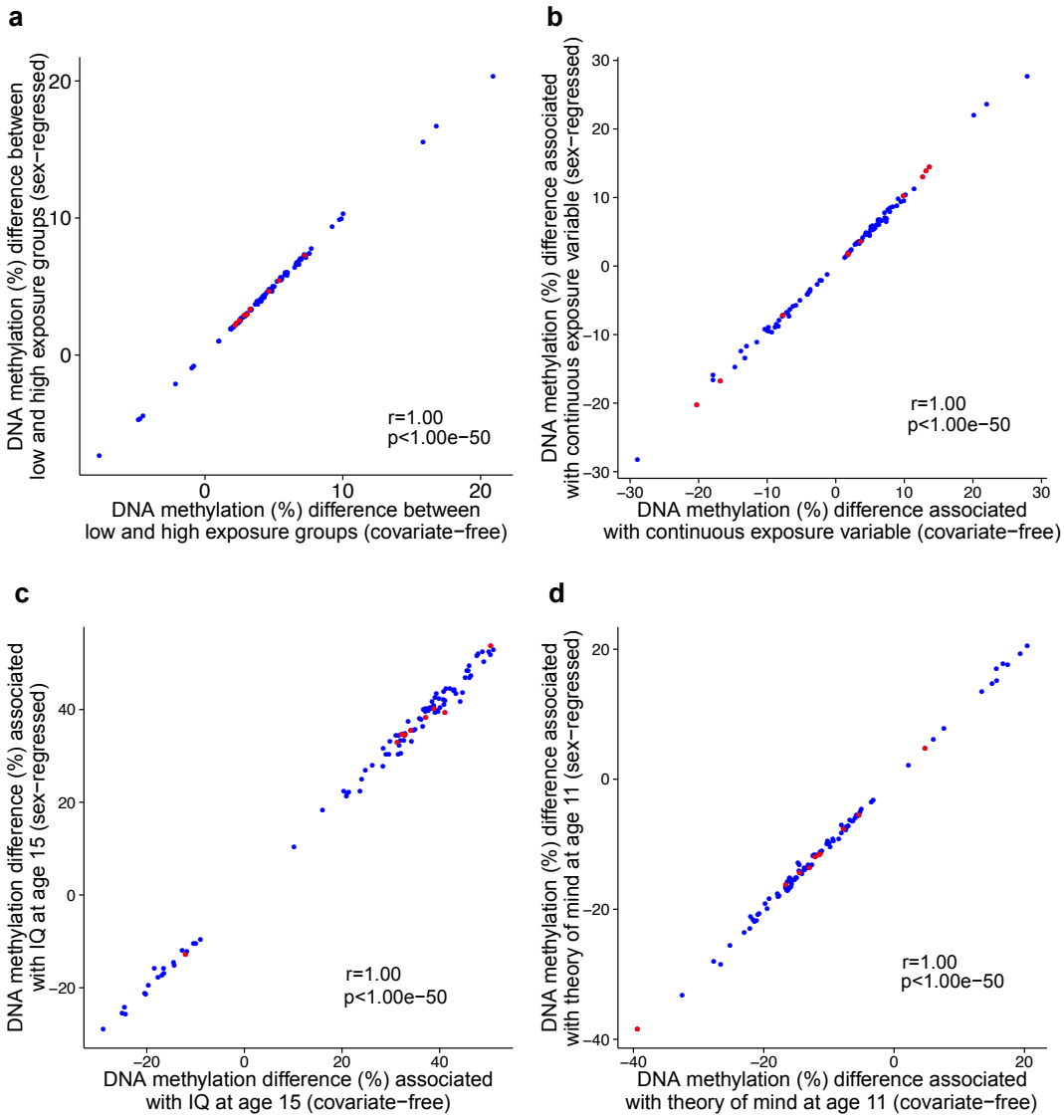


Figure 2-4. Associations of DNA methylation with exposure and outcome measures were highly correlated comparing sex-regressed and covariate free models. Effect sizes of the 100 top-ranked DMPs associated with (**a**) exposure group (> 6 months vs < 6 months) (**b**) quantitative measures of exposure time (**c**) IQ at age 15 and (**d**) Theory of Mind at age 11 in covariate-free tests are correlated almost perfectly effect sizes in sex-regressed models at the same 100 probes (r = 1.00, $P \leq 1.00 \times 10^{-50}$ for each test). The top ten associated sites for the covariate-free model are highlighted in red in each of the panels

Supplementary Figure 2

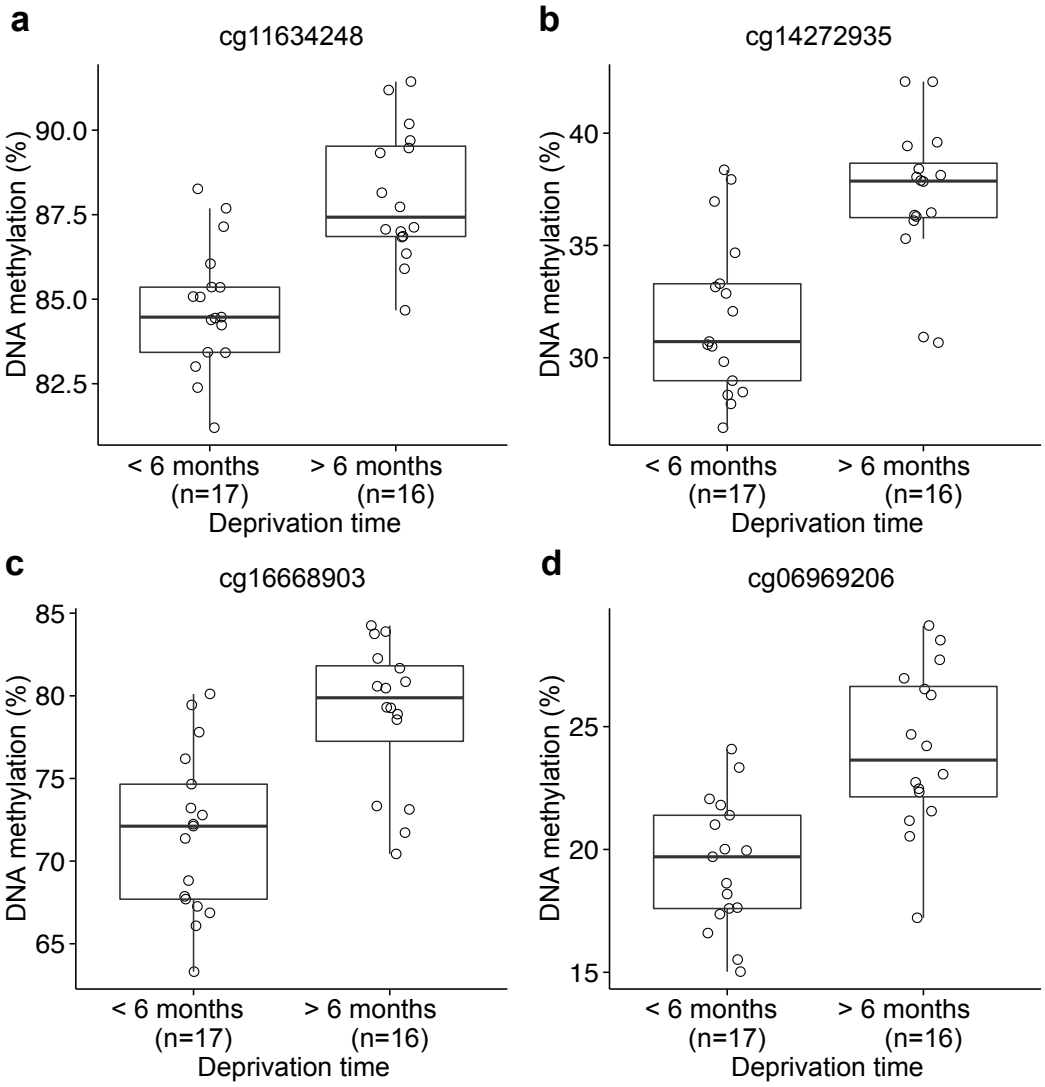


Figure 2-5. The four top-ranked DMPs associated with severe early-life adversity. Associations were identified comparing the >6 months and <6 months exposure groups. (**a**) cg11634248, $P = 2.35 \times 10^{-5}$, (**b**) cg14272935, $P = 6.89 \times 10^{-5}$, (**c**) cg16668903, $P = 9.75 \times 10^{-5}$, (**d**) cg06969206, $P = 1.08 \times 10^{-4}$.

Supplementary Figure 3

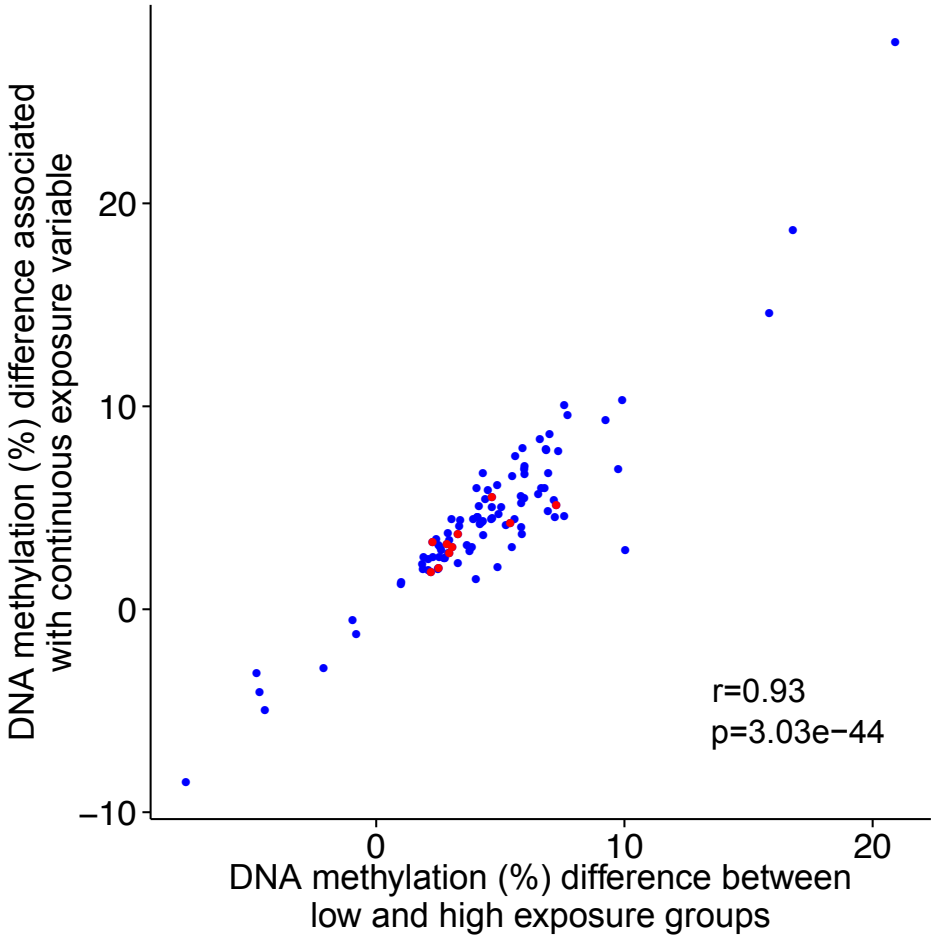


Figure 2-6. Correlations between the top 100 exposure-associated DMPs and continuous exposure duration. For the 100 top ranked exposure-associated DMPs (see **Supplementary Table 4**) effect sizes for association with exposure group correlated significantly with effect sizes for association with continuous exposure duration (r = 0.93, P = 3.03 × 10^{-44}). The top ten DMPs associated with exposure group (see **Table 2-1**) are highlighted in red.

Supplementary Figure 4

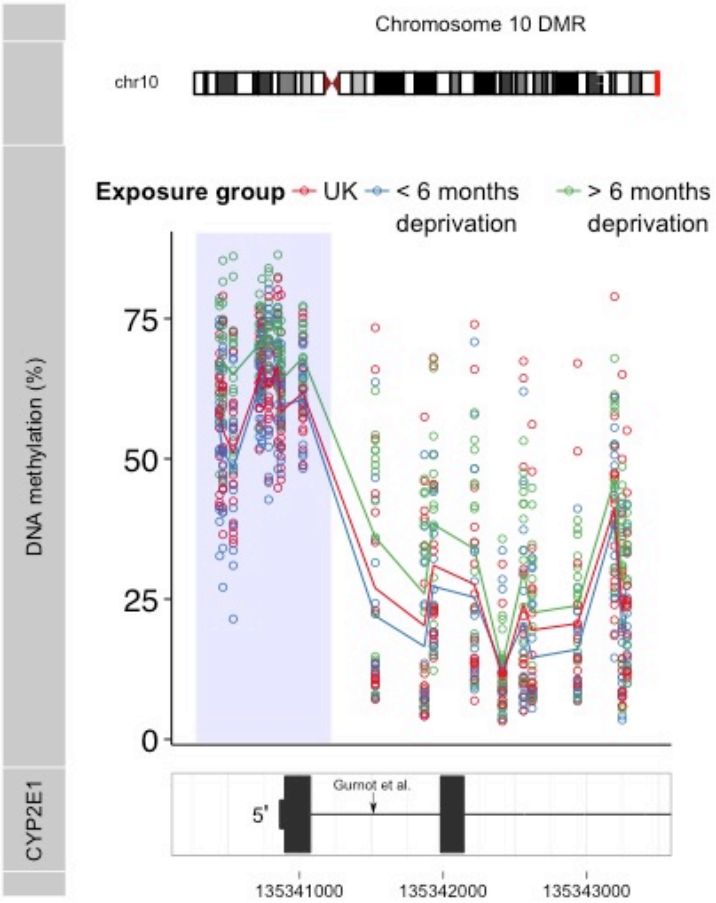


Figure 2-7. A DMR on chromosome 10 spanning nine sequential 450K array probes (chr10:135340445-135341026) was identified by *comb-P.* DNA methylation across this region is significantly elevated ($P = 2.21 \times 10^{-10}$; corrected Šidák $P = 2.98 \times 10^{-5}$) in individuals exposed to severe institutional deprivation. The >6 months institutionalized group (green) is characterized by consistent hypermethylation across the whole DMR compared to the <6 months institutionalized group (blue) and UK control group (red). An extended ~3kb region spanning the first two exons of *CYP2E1* is shown with the DMR region highlighted in light blue. The probe found to be associated with prenatal maternal antidepressant exposure in neonates by Gurnot *et al* (2015) is highlighted.

Supplementary Figure 5

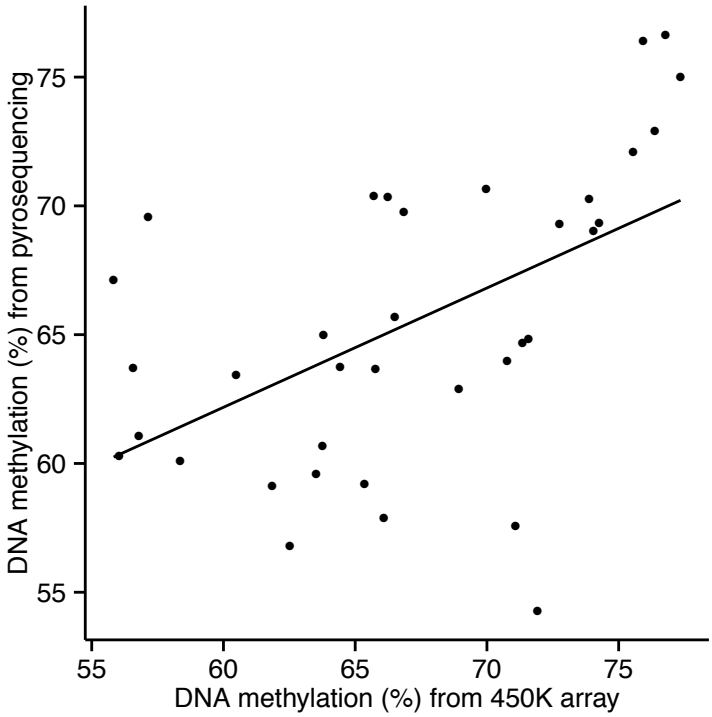


Figure 2-8. DNA methylation values for CpG sites within the *CYP2E1* DMR quantified using the Illumina 450K array were validated using bisulfite-pyrosequencing. The assay spanned three CpG sites (cg14250048, cg00436603, and cg01465364) within the DMR, for which there was a highly-significant correlation between DNA methylation levels independently derived from the 450K array and bisulfite-pyrosequencing experiments (r = 0.52, P = 0.001). Shown is the average DNA methylation (%) across the three CpG sites as measured by bisulfite-pyrosequencing and the 450K array for each of the 36 individuals profiled by both methodologies.

Supplementary Figure 6

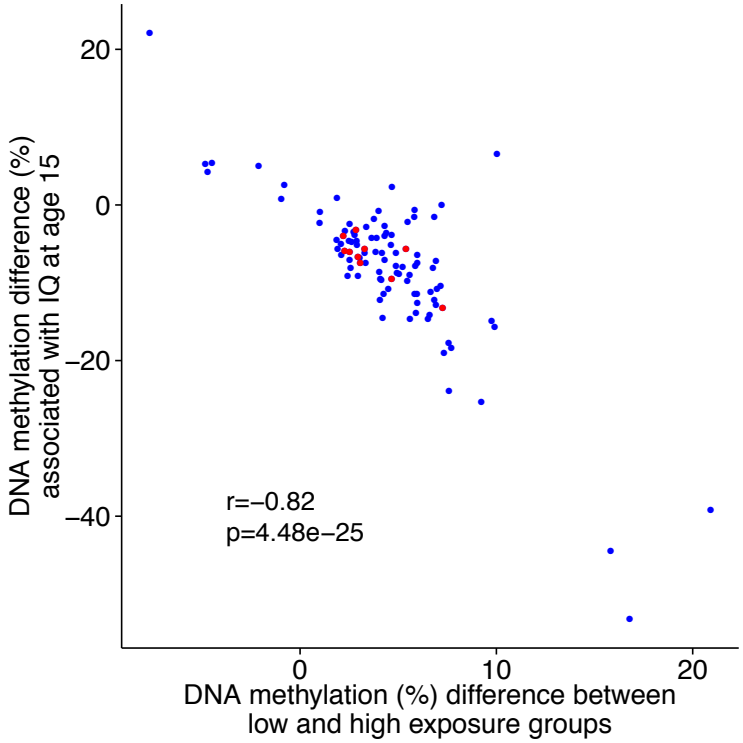


Figure 2-9. Correlations between the top 100 exposure-associated DMPs and IQ at age 15. For the 100 top ranked exposure-associated DMPs (see **Table 2-7**) effect sizes for association with exposure group correlated significantly with effect sizes for association with IQ at age 15 (r = -0.82, $P = 4.48 \times 10^{-25}$). The top ten DMPs associated with exposure group (see **Table 2-1**) are highlighted in red.

3. Comprehensive DNA-methylation profiling of victimised young people reveals no association between stress and epigenetic variation in blood

3.1 Introduction

The question of "how stress gets under the skin" has intrigued and puzzled researchers for decades (Rutter, 1981, Lupien et al., 2009, McEwen, 2012). Early-life stress and victimisation have been linked to numerous and often persistent adverse health outcomes, both physical and psychological (Fisher et al., 2015, Gilbert et al., 2009, Widom et al., 2012, Anda et al., 2006), including but not limited to obesity (Baldwin et al., 2016), inflammation (Danese et al., 2009) and depression (Danese et al., 2009). Furthermore, these adverse exposures have been associated with neurobiological sequelae including altered brain development, plasticity and structure (McEwen, 2012, Lupien et al., 2009, Teicher and Samson, 2016). Direct effects of stress on the neuroendocrine response system involving HPA axis activation and glucocorticoid release have been identified both in animal (Liu et al., 1997) and human work (Taylor, 2010, Gunnar and Quevedo, 2007). Epigenetic modifications and specifically DNA methylation have been proposed as the missing link in the cascade between stress exposure and outcomes, providing a heritable but dynamically adaptive mechanism by which the organism responds to the environment and resulting in stable gene expression changes (Petronis, 2010, Szyf and Bick, 2013, Gluckman et al., 2008, Provencal and Binder, 2015, Lutz and Turecki, 2014).

An increasing body of research has experimentally linked adverse early-life environments to alterations in DNA methylation in brain tissue of rodents (Weaver et al., 2004, Desarnaud et al., 2008, Kember et al., 2012, Kundakovic et al., 2013, Doherty et al., 2016, Saunderson et al., 2016, Murgatroyd et al., 2009, Roth et al., 2009) and primates (Provencal et al., 2012). There appears to be a clear focus on genes involved in HPA axis response with a majority of the work concentrating on the glucocorticoid receptor gene *NR3C1*, and some studies considering *AVP* (Kember et al., 2012, Murgatroyd et al., 2009) and *BDNF* (Kosten and Nielsen, 2014, Doherty et al., 2016, Roth et al., 2009). Most studies in human samples have been observational and often conducted in surrogate tissue, including blood, epithelial tissue or saliva. A substantial body of research has accumulated linking early-life stress to differences in DNA methylation in the glucocorticoid receptor (Palma-Gudiel et al., 2015). A few studies have been able to conduct research on post-mortem human brain tissue, confirming the differences in DNA methylation at *NR3C1* observed in rodent brains for human

samples (Labonte et al., 2012, Suderman et al., 2012, McGowan et al., 2009). Studies in peripheral tissues have furthermore linked methylation in a variety of other genes to adverse early-life environments, including *FKBP5* (Tyrka et al., 2012), *SLC6A4* (Ouellet-Morin et al., 2013, Booij et al., 2015, Devlin et al., 2010), and *KITLG* (Houtepen et al., 2016). The only quasi-experimental settings, in which exposure to adverse early-life environments is randomised, involve studies of institutionalisation (Naumova et al., 2012, Non et al., 2016, Kumsta et al., 2016). Findings identifying differential DNA methylation associated with the institutionalisation exposure from these studies are sparse and non-overlapping, but include differential methylation at the genes *FKBP5*, *SLC6A4* and *CYP2E1*.

Despite the large number of studies, few have taken an unbiased, genome-wide approach and there appears to be a mostly heterogeneous picture when considering results from this subset of studies (Houtepen et al., 2016, Kumsta et al., 2016, Weder et al., 2014). In addition to the results the exposure variables are varied and include institutionalisation (Naumova et al., 2012), maternal psychiatric disease (Devlin et al., 2010), sexual abuse (Beach et al., 2010), neglect (Weder et al., 2014), and trauma (Tyrka et al., 2012); many of which were collected retrospectively and are potentially liable to confounding (Danese et al., 2016). Other types of confounding including from familial environments, toxic substance exposure or cell-type composition in studies using whole blood are rarely addressed. Most studies in animal and human samples are characterised by very small sample sizes, even though most research on human samples was conducted in peripheral tissues.

In this study, we undertook the largest and most systematic investigation of early-life victimisation and its association with DNA methylation in whole blood to date (n = 1,658). Using the Environmental Risk (E-Risk) Longitudinal Twin Study we tested whether victimisation exposure in adolescence or childhood was associated with altered DNA methylation in whole blood at age 18, profiled on the Illumina 450K HumanMethylation Array. To interrogate the effects of specific forms of victimisation, we ascertained multiple different victimisation types across the first two decades of life, allowing us to study the combined, shared and unique effects of these exposures. We were able to examine two major developmental periods, adolescence and childhood, teasing apart their individual and cumulative epigenomic correlates. Our full sample of 1,658 individuals profiled for DNA

methylation allowed us to undertake the most powerful epigenomic association study of early-life victimisation to date, at the same time we exploited the twin design of the E-Risk study to investigate twin pair epigenomic differences linked to divergent victimisation exposures, enabling us to control for familial environments as well as genetic effects. Our main aim was to provide a comprehensive and systematic analysis of DNA methylation signatures of early-life victimisation in blood and their potential role in the biological embedding of developmental adversities.

3.2 Methods

3.2.1 Study sample

Participants were members of the Environmental Risk (E-Risk) Longitudinal Twin Study, which tracks the development of a birth cohort of 2,232 British children. The sample was drawn from a larger birth register of twins born in England and Wales in 1994-95 (Trouton et al., 2002). Full details about the sample were reported previously (Moffitt and E-Risk Study Team, 2002). Briefly, the E-Risk sample was established in 1999-2000, when 1,116 families (93% of those eligible) with same-sex 5-year-old twins participated in home-visit assessments. This sample consisted of 55% monozygotic (MZ) and 45% dizygotic (DZ) twin pairs; sex was evenly distributed within zygosity (49% male) and 7% of the sample self-reported as Black, Asian, or mixed-race. Families were recruited to be representative of the UK population with newborns in the 1990s, to ensure adequate numbers of children in disadvantaged homes and to avoid an excess of twins born to well-educated women using assisted reproduction. The study sample represents the full range of socioeconomic conditions in Great Britain, as reflected in the families' distribution on a neighbourhood-level socioeconomic index (Odgers et al., 2012).

Follow-up home visits were conducted when children were aged 7 (98% participation), 10 (96% participation), 12 (96% participation), and most recently, 18 years (93% participation). At age 18, 2,066 participants were assessed. The average age at the time of assessment was 18.4 years (SD = 0.36); all interviews were conducted after the 18th birthday.

Our epigenetic study used DNA from a single tissue: blood. At age 18, whole

blood was collected from 82% (n = 1,700) of the participants in 10mL K_2EDTA tubes. DNA was extracted from the buffy coat using a Flexigene DNA extraction kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. There were no differences between those who did and did not provide blood in terms of SES (*P* = 0.34, chi-squared test), age-5 IQ scores (*P* = 0.14, t-test), age-5 internalising or externalising behaviour problems (*P* = 0.23 and *P* = 0.64, t-tests), childhood polyvictimisation (*P* = 0.42, Z-test), or adolescent polyvictimisation (*P* = 0.21, Z-test). Study members who did not provide blood provided buccal swabs, but these were not included in our methylation analysis to avoid tissue-source confounds.

The Joint South London and Maudsley and the Institute of Psychiatry Research Ethics Committee approved each phase of the study. Parents gave informed consent and twins gave assent between 5-12 years and then informed consent at age 18. The analysis plan for this study was posted in advance (http://www.moffittcaspi.com).

3.2.2 Measures

Adolescent victimisation. The adolescent victimisation measures have been described previously (Fisher et al., 2015). Briefly, at age 18, participants were interviewed about exposure to a range of adverse experiences between 12-18 years using the Juvenile Victimization Questionnaire (JVQ) (Finkelhor et al., 2011, Hamby et al., 2004), adapted as a clinical interview. Our adapted JVQ comprised 45 questions covering seven different forms of victimisation: Crime victimisation, peer/sibling victimisation, internet/mobile phone victimisation, sexual victimisation, family violence, maltreatment, and neglect. Exposure to each type of adolescent victimisation was coded on a 5-point scale, in which "0" indicated "no exposure" and "5" indicated "definite" or "very severe" exposure.

The adolescent polyvictimisation variable was derived by summing all victimisation experiences that received a code of at least "4": (i.e., severe exposure). We winsorized the polyvictimisation distribution into a four-category variable (representing 0, 1, 2, 3+ experiences), as we did childhood polyvictimisation (**Table 3-1**). An overview of all victimisation exposures analysed in adolescence is given in **Table 3-2**.

Table 3-1. Overview of adolescent victimisation, smoking exposure and SES. Shown are the distributions of gender, smoking exposure and socioeconomic status (SES) across adolescent polyvictimisation categories in sample profiled for DNA methylation.

Adolescent polyvictimisation	0 (n=1064)	1 (n=325)	2 (n=150)	3+ (n=118)	NA (n=1)
m/f	538/526	171/154	74/76	48/70	1/0
Smoking yes/no	155/908	98/226	51/99	73/45	0/1
Pack-years (mean)	0.28	0.77	0.87	1.75	-
SES (mean)	2.10	1.89	1.83	1.69	-
Childhood polyvictimisation	0 (n=1192)	1 (n=355)	2 (n=70)	3+ (n=41)	
m/f	594/598	169/186	41/29	28/13	
Smoking yes/no	207/983	111/244	38/32	21/20	
Pack-years (mean)	0.37	0.75	1.45	1.85	
SES (mean)	2.13	1.77	1.59	1.24	
Cumulative polyvictimisation	0 (n=834)	1 (n=254)	2 (n=412)	3 (n=158)	
m/f	425/409	138/116	190/222	79/79	
Smoking yes/no	102/731	65/189	125/286	85/73	
Pack-years (mean)	0.22	0.59	0.69	1.68	
SES (mean)	2.19	1.67	2.03	1.51	

Childhood victimisation. The childhood victimisation measures have been described previously (Danese et al., 2016). Briefly, exposure to several types of victimisation was assessed repeatedly when the children were 5, 7, 10, and 12 years of age. These were exposure to domestic violence between the mother and her partner; frequent bullying by peers; physical maltreatment by an adult; sexual abuse; emotional abuse; and physical neglect. Exposure to each type of victimisation was coded on a 3-point scale, in which “0” indicated “no exposure,” “1” indicated “probable” or “less severe” exposure, and “2” indicated “definite” or “severe” exposure.

We operationalised polyvictimisation as the count of forms of victimisation experienced by a child. This variable was derived by summing all childhood victimisation experiences coded as 2. Next, we winsorized the polyvictimisation distribution into a four-category variable (representing 0, 1, 2, 3+ experiences; **Table 3-1**). An overview of all victimisation exposures analysed in childhood is given in **Table 3-2**.

Cumulative victimisation. Latent Class Analysis (LCA) is a person-centered analytical approach that classifies individuals into groups or classes based on a profile of variables, in this case the degree of exposure (i.e., none, moderate, or severe) to the six types of victimisation experienced during childhood (ages 5-12 years) and seven types of victimisation experienced during adolescence (ages 12-18 years). Since we were interested in the profiles of individuals who were victimised, we only classified n = 1,043 subject members who experienced at least one severe form of victimisation across childhood or adolescence. LCA was conducted in MPlus v7.4 and accounted for clustering of twins within families. The results of the Latent Class Analysis are described in the Results section and **Table 3-1**. For sensitivity analyses we additionally operationalised cumulative victimization as the sum of the childhood and adolescent victimization (**Table 3-2**).

Smoking. We assessed smoking among E-risk participants by calculating pack-years, where pack-years = (number of cigarettes smoked per day × number of years smoked) / 20; one pack-year = 7,305 cigarettes.

Socioeconomic status. This measure has been described previously (Trzesniewski et al., 2006): socioeconomic status (SES) was operationalised as a standardised composite of income, education, and occupation.

3.2.3 Genome-wide quantification of DNA methylation

Experimental procedure. We assayed 1,669 samples (out of 1,700 available whole blood samples); 31 samples were not useable (e.g., due to low DNA concentration). ~500ng of DNA from each sample was treated with sodium bisulfite, using the EZ-96 DNA Methylation kit (Zymo Research, CA, USA). DNA methylation was quantified using the Illumina Infinium HumanMethylation450 BeadChip (Illumina Inc, CA, USA) ("Illumina 450K array") run on an Illumina iScan System (Illumina, CA, USA) using the manufacturers' standard protocol.

Members of complete twin pairs (n = 743 pairs; n = 1,486 individuals) were randomly assigned to bisulfite-conversion plates and Illumina 450K arrays, with siblings processed in adjacent positions to minimise batch effects. Singletons (n = 183) – i.e., members of incomplete twin-pairs - were randomised and processed on two separate bisulfite-conversion plates. Fully methylated control samples (CpG Methylated HeLa Genomic DNA; New England BioLabs, MA, USA) were included in a random position on each plate; the distinct DNA methylation profile of this sample enabled us to confirm the experiment was successful and to ensure there were no plate mix-ups or rotations.

Data pre-processing. Data from 1,669 unique individuals entered our Quality Control (QC) pipeline, which was performed in the R statistical programing environment (R Core Team, 2013). Data were imported using the *methylumIDAT()* function from the *methylumi* package (Davis et al., 2015). The first QC step assessed the signal intensities, excluding samples with both median methylated ('M') and unmethylated ('U') intensities < 2500. Second, the fully methylated controls were identified - their intensity profiles were characteristic of the sample being fully methylated and confirmed that no plate rotations or plate mislabelling had occurred – and subsequently removed from the dataset. Third, using ten control probes included on the 450K array, we examined the efficiency of the sodium bisulfite conversion reaction; samples were excluded if their "conversion score" was < 80. Fourth, multidimensional scaling was performed for DNA methylation probes on each of the sex chromosomes and compared to the reported gender. Fifth, to confirm genetic identity of the DNA samples, we assessed genotype concordance between SNP probes on the 450K array and data generated using Illumina OmniExpress24v1.2 genotyping BeadChips.

1,658 samples passed our stringent QC pipeline, which will be referred to as the 'Sample of Individuals'. The data were then processed with the *pfilter*() function from the *wateRmelon* package (Pidsley et al., 2013) excluding 0 samples with > 1% of sites with a detection *P* value > 0.05, 567 sites with bead count < 3 in 5% of samples and 1,448 probes with > 1% of samples with detection *P* value > 0.05. The data were normalised with the *dasen*() function from the *wateRmelon* package (Pidsley et al., 2013). From this dataset, we extracted all complete twin pairs (i.e. both members passed QC and were run on the same chip) (n = 732 pairs; n = 1,464 individuals). This will be referred to as the 'Sample of Twin Pairs'.

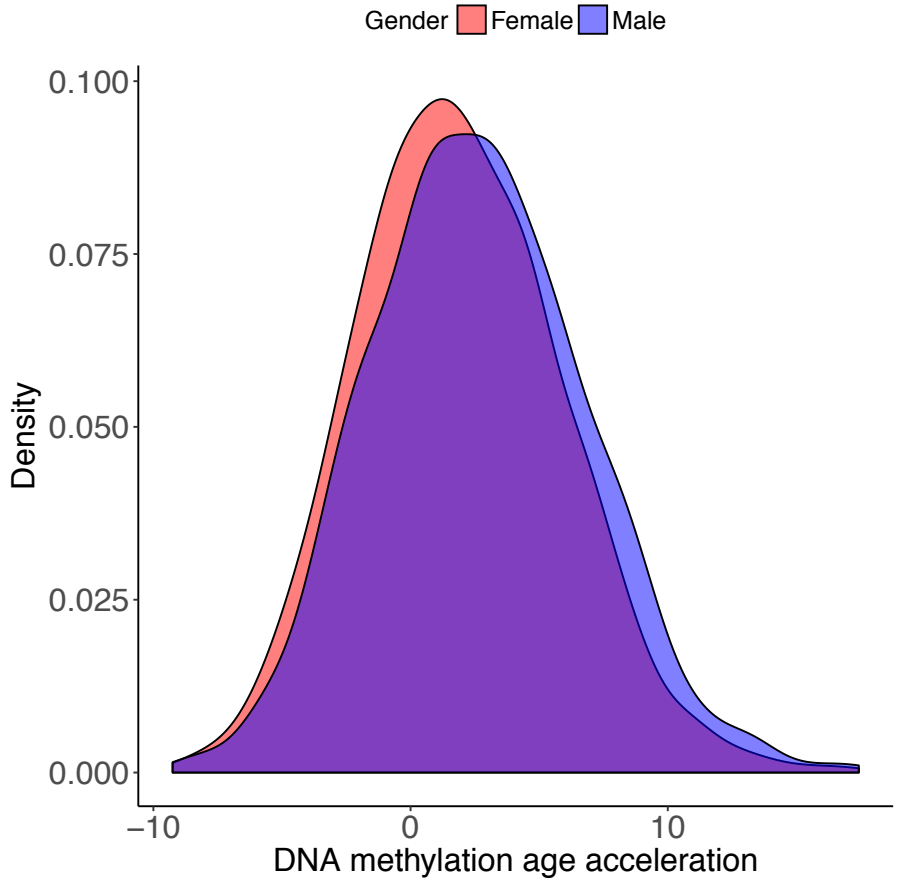


Figure 3-1. Distribution of DNA methylation age acceleration in the E-Risk EWAS sample.

As a further check, DNA methylation values for all samples were uploaded to the online Epigenetic Clock software (Horvath, 2013) for DNA methylation age (DNAmAge) and tissue prediction. The majority of samples (72%) were predicted within 5 years as illustrated by the distribution of DNA methylation age acceleration (DNA methylation age acceleration = chronological age – DNA methylation age estimate; **Figure 3-1**) and all were predicted as of blood origin. Prior to any analyses, probes with common (> 5% MAF) SNPs within 10bp of the single base extension and probes with sequences previously identified as potentially hybridising to multiple genomic loci were excluded (Price et al., 2013, Chen et al., 2013).

3.2.4 Statistical methods

All statistical analyses were conducted in R version 3.3.0 unless otherwise specified.

Testing the association between victimisation and DNA methylation: An individual-level analysis

Our first analysis used the Sample of Individuals, which includes all E-risk participants included in DNA methylation profiling who passed QC and filtering. We used a linear model to test the association between DNA methylation variation (dependent variable) and victimisation. The model included the following covariates: Sex, batch, and cell-type proportion estimates. Because the sample included members of twin pairs, we accounted for the non-independence of observations by calculating robust standard errors using the R packages *plm* (Croissant and Millo, 2008), *lme4* (Bates et al., 2015) and *sandwich* (Zeileis, 2004). To control for smoking, we re-estimated the model by adding smoking pack-years (see **Measures**) as an additional covariate.

The same basic linear model was used to test methylomic variation associated with (1) adolescent polyvictimisation and its seven subtypes of victimisation: maltreatment, neglect, sexual victimisation, family violence, peer/sibling victimisation, internet/mobile phone victimisation and crime victimisation; (2) childhood polyvictimisation and its six subtypes: physical abuse, physical neglect, emotional neglect, sexual victimisation, domestic violence and bullying; and (3) cumulative polyvictimisation. The victimisation variables are summarised in **Table 3-2**.

Testing the association between victimisation and DNA methylation: A twin-difference analysis

Because the majority (88%) of individuals in E-Risk who were profiled for DNA methylation are members of a twin-pair, we were able to use a discordant-twin design to control for the confounding effects of family background and genetic factors when testing the association between victimisation and DNA methylation. We ran a dizygotic (DZ) twin-difference model to control for family background and a monozygotic (MZ) twin-difference model to correct for both family

background and genetic influences for each of the three main exposure measures: (1) adolescent polyvictimisation, (2) childhood polyvictimisation, and (3) cumulative polyvictimisation. A linear fixed effects model was used to test the association between DNA methylation differences within twin pairs (dependent variable) and differences in their victimisation exposure, controlling for differences in blood cell-type proportion estimates. Because members of each twin pair were run in the same batch and are of the same sex, controlling for batch and sex was not necessary.

Table 3-2. Victimisation variables used in this study

Time period	Victimisation type	Variable measurement	Levels	Type
Adolescence	Polyvictimisation - continuous	Sum of individual victimisation types, winsorized at 3.	0-3+	Continuous
	Maltreatment	Severity code	0-5	Continuous
	Neglect	Severity code	0-5	Continuous
	Sexual victimisation	Severity code	0-5	Continuous
	Family violence	Severity code	0-5	Continuous
	Peer/sibling victimisation	Severity code.	0-5	Continuous
	Internet/mobile phone victimisation	Severity code	0-5	Continuous
	Crime victimisation	Severity code	0-5	Continuous
Childhood	Polyvictimisation - continuous	Sum of individual victimisation types, winsorized at 3.	0-3+	Continuous
	Physical abuse	Severity code	0-2	Continuous
	Physical neglect	Severity code	0-2	Continuous
	Emotional neglect	Severity code	0-2	Continuous
	Sexual victimisation	Severity code	0-2	Continuous
	Domestic violence	Severity code.	0-2	Continuous
	Bullying	Severity code	0-2	Continuous

Cumulative	Latent Class Analysis (LCA) - factor	A latent class analysis of individual childhood and adolescent victimisation types resulted in four subgroups of participants: 0. never victimised, 1. domestic violence during childhood, 2. victimised by peers and 'street' crime, 3. revictimised by multiple severe types of victimisation	0-3	Factor
	Additive model	Cumulative victimisation = Childhood + Adolescent polyvictimisation	0-6	Continuous

Multiple testing threshold

To establish an array-wide multiple testing significance threshold, 5,000 permutations were performed repeating a linear regression model for randomly selected groups of cases and controls. For each permutation, *P* values from the EWAS were saved and the minimum identified. Across all permutations the 5th percentile was calculated to generate the 5% alpha significance threshold ($P < 1.61 \times 10^{-7}$). Identifying a study-wide multiple-testing threshold would be extremely challenging, given the correlated nature of the analyses performed on each phenotype (robust standard error EWAS, twin difference models, smoking correction) as well as the correlation of the different victimisation exposure measures. Therefore, we report raw *P* values for all analyses, highlighting the array-wide significance level in plots and text.

Cell counts

Cell-type heterogeneity is known to affect studies of DNA methylation and should therefore be adjusted for where possible (Houseman et al., 2012). Cell-type proportion estimates were obtained using the online tool at https://dnamage.genetics.ucla.edu (Horvath, 2013), which bases its estimates on the Houseman et al reference-based estimator (Houseman et al., 2012) for CD4+ T cells, natural killer cells, monocytes and granulocytes and on an algorithm developed by Horvath (Horvath, 2013) for naïve CD8 T cells, CD8+ T cells and plasmablasts (**Figure 3-2**).

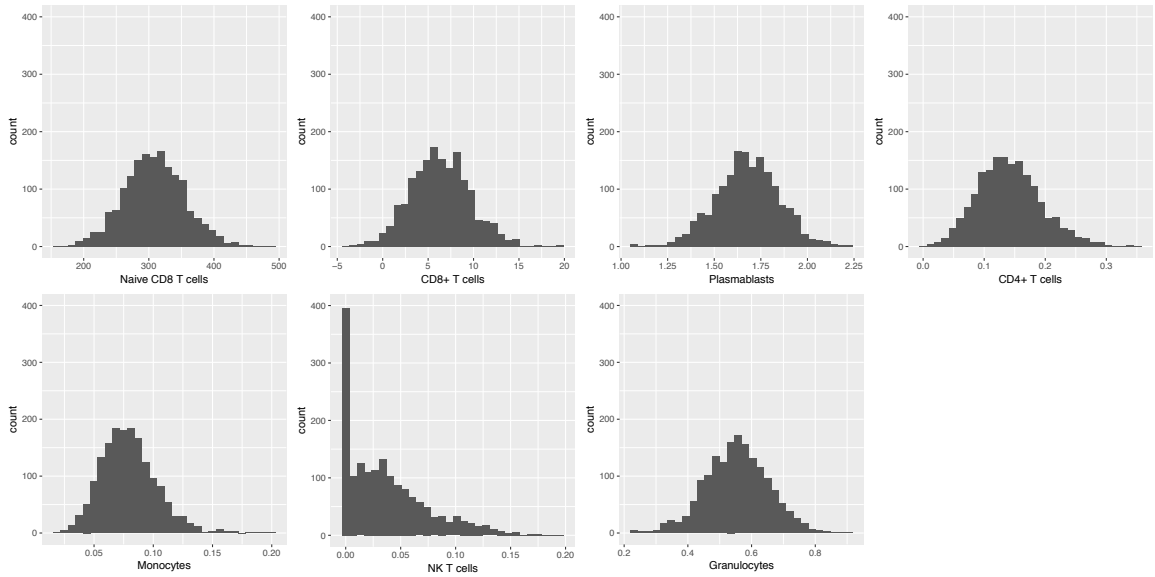


Figure 3-2. Distributions of blood cell-type estimates in the E-Risk EWAS sample.

Candidate gene analyses

We interrogated candidate genes which are known to be involved in stress reactivity and which have been previously reported to be associated with response to trauma and with indices of psychosocial stress. The candidates comprise HPA-axis genes *N3RC1*, *FKBP5*, *CRHR1*, *BDNF* and *AVP*. Additionally, we included *SLC6A4* because of its link to stress reactivity (Caspi et al., 2003) as well as *CYP2E1*, which has been shown to be differentially methylated in the context of severe early-life institutional deprivation (Kumsta et al., 2016). Probes were annotated to genes using the *Genomic Region Enrichment and Annotation Tools* (*GREAT*) (McLean et al., 2010). For each candidate gene, we examined the association between all annotated probes and adolescent and childhood victimisation, respectively. We provide a regional Manhattan plot showing the array-wide *P* value threshold as well as a gene-wide significance threshold based on a Bonferroni-correction for the number of probes annotated to the gene. In addition, we calculated combined gene-wide *P* values using Brown's method (Brown, 1975), which are reported in **Appendix A**.

Regional Manhattan plots were generated using the Bioconductor package *ggbio* (Yin et al., 2012).

Testing the association between victimisation and epigenetic age acceleration

For each participant, we calculated DNA methylation age acceleration by subtracting their chronological age at the time blood was collected from their DNA methylation age estimate based on the epigenetic clock (Horvath, 2013). DNA methylation age acceleration was regressed onto adolescent, childhood and cumulative victimisation, respectively, using robust standard errors as described above.

3.3 Results

3.3.1 Victimisation in the peak period of adolescence

Adolescents are victimised by a more diverse set of actors across a wider range of environments than any other age group. Exposure to multiple types of victimisation - including relational aggression, sexual victimisation, and serious violent crime - peaks during adolescence (Brown et al., 2005, Sickmund and Puzzanchera, 2014, Peskin et al., 2006). Therefore, we tested whether victimisation during this peak period of victimisation exposure was associated with methylomic variation quantified in whole blood at age 18. We regressed DNA methylation onto adolescent polyvictimisation (see **Methods, Table 3-1**) accounting for family structure across all 1,658 samples passing our stringent QC pipeline (see **Methods)**, and including batch, sex and derived blood cell-type proportion estimates as covariates. DNA methylation at four differentially methylated positions (DMPs) (cg05575921, cg26703534 and cg21161138 annotated to *AHRR* and *C5orf55*, and cg14179389 annotated to *GLMN* and *GFI1*) was associated with adolescent polyvictimisation at an array-wide significance level ($P < 1.61 \times 10^{-7}$; **Figure 3-3**, top panel; **Table 3-3**). Across all probes on the array, the effect sizes were small (**Figure 3-4**) and each of the four significant DMPs were characterised by methylation differences of less than 1.5% for each additional type of victimisation (**Figure 3-5**).

Table 3-3. We identified four array-wide significant DMPs associated with adolescent polyvictimisation. Shown for each DMP are chromosomal location (hg19), annotated genes as identified by *GREAT (McLean et al., 2010)*, *P* value from our linear model with robust standard errors (**Methods**) and effect size (difference in DNA methylation (%) for each additional form of victimisation experienced).

Probe	Position	Annotated genes	*P* value	Effect size
cg05575921	5:373378	*AHRR; C5orf55*	2.20×10^{-10}	-1.36
cg26703534	5:377358	*C5orf55; AHRR*	2.57×10^{-10}	-0.67
cg21161138	5:399360	*C5orf55; AHRR*	1.52×10^{-08}	-0.67
cg14179389	1:92947961	*GLMN; GFI1*	1.65×10^{-08}	-0.89

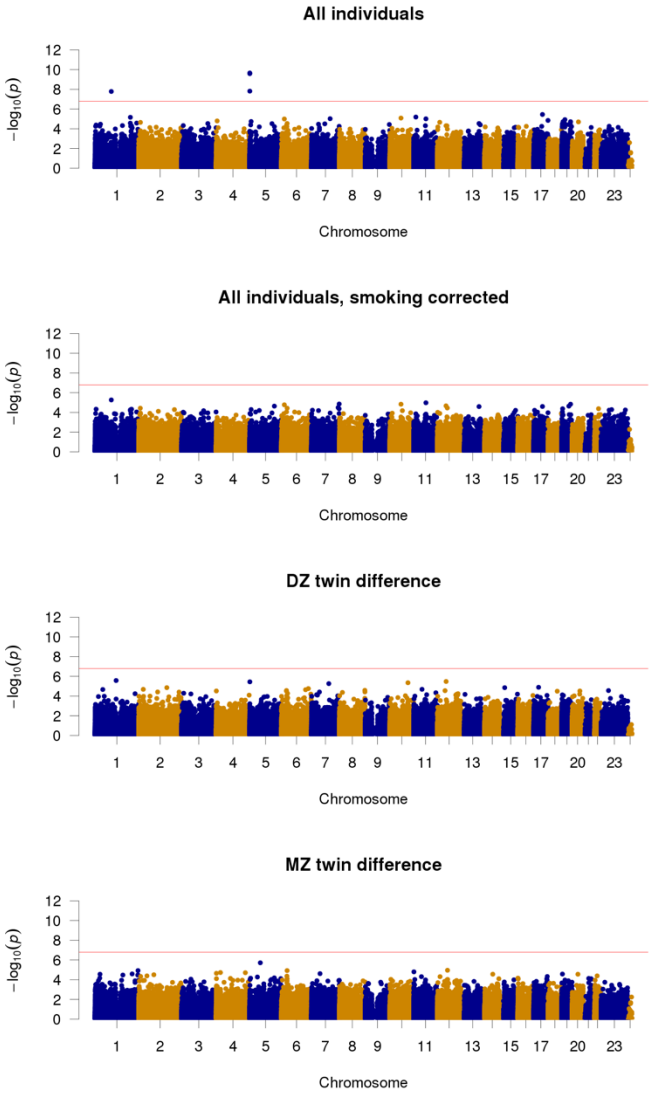


Figure 3-3. The association between DNA methylation and adolescent polyvictimisation is confounded by smoking. The four panels show $-\log_{10} P$ values for the association between DNA methylation and adolescent polyvictimisation. **Top panel**: Four probes pass the array-wide multiple testing threshold ($P < 1.61 \times 10^{-7}$) in a regression model controlling for batch, sex and cell-type proportion estimates (see **Table 3-3**). **Second panel**: We identify no significant DMPs when adding smoking pack-years as a further covariate. The bottom panels show *P* value distributions from two twin difference models, which examine the association between twin-pair differences in DNA methylation and differences in polyvictimisation, while controlling for differences in cell-type proportion estimates. No significant associations were observed for either the DZ (**third panel**) or MZ twin difference model (**fourth panel**).

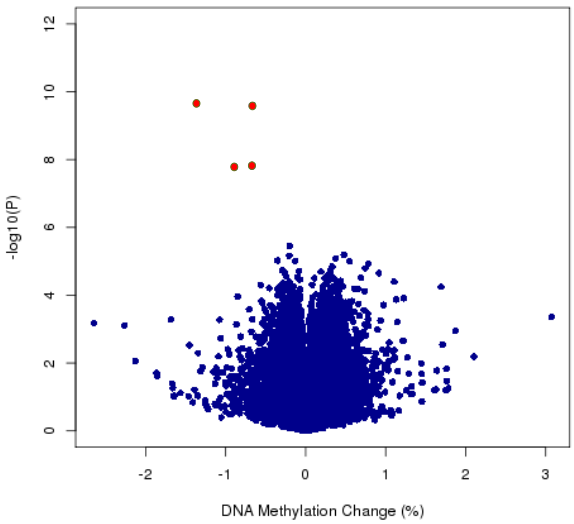


Figure 3-4. Volcano plot showing effect sizes and *P* values for the association between DNA methylation and adolescent polyvictimisation. The distribution of $-\log_{10}$ *P* values and effect sizes (DNA methylation change for each increase in polyvictimisation score) are shown. The four array-wide significant probes are highlighted in red.

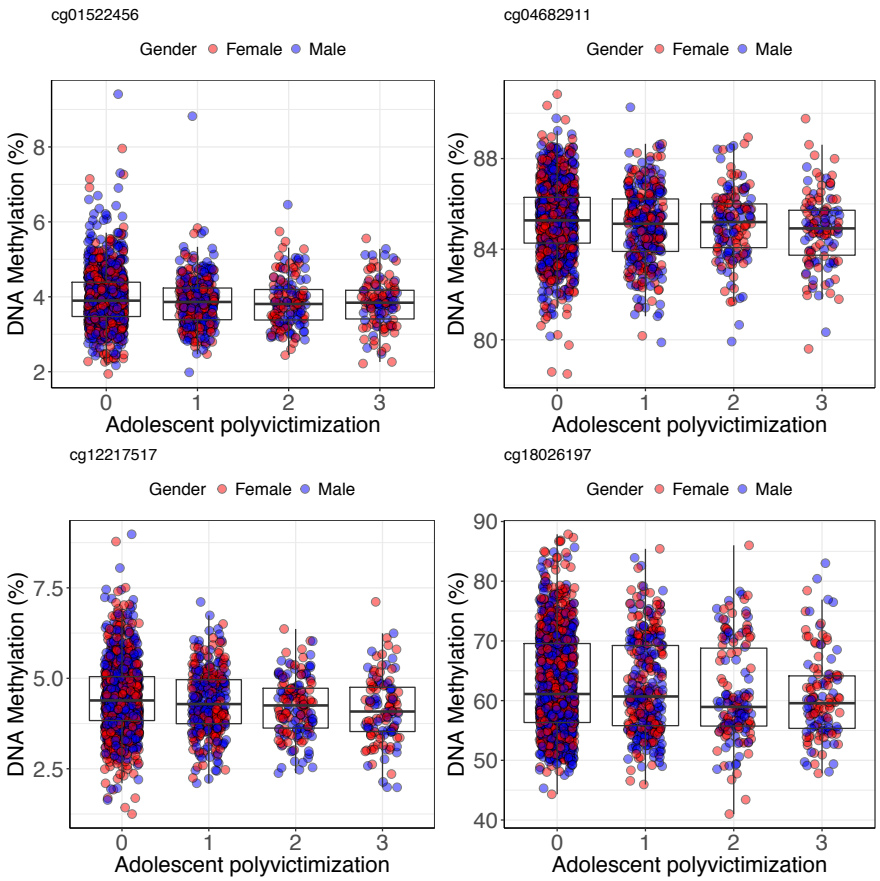


Figure 3-5. DNA methylation patterns across the four array-wide significant probes associated with adolescent polyvictimisation.

One important challenge to interpreting the association between victimisation and DNA methylation is that victimisation is often correlated with exposure to environmental toxins; specifically, victimised adolescents are more likely to smoke (Nichols and Harlow, 2004, Norman et al., 2012). Additionally, exposure to tobacco smoke has been reported to be associated with striking effects on DNA methylation at specific CpG sites across the genome (Elliott et al., 2014, Zeilinger et al., 2013, Tsaprouni et al., 2014) and is an important confounder in epigenetic epidemiology (Hannon et al., 2016a). In particular, DNA methylation differences in the gene *AHRR*, to which three of our four array-wide significant probes are annotated, have been found to be robustly associated with smoking (Joehanes et al., 2016). We assessed smoking among E-Risk participants by calculating the number of pack-years that they smoked (see **Methods**), observing that victimised adolescents were significantly more likely to smoke ($P = 2.51 \times 10^{-37}$), and to smoke more pack-years ($P = 4.92 \times 10^{-14}$, **Figure 3-6**).

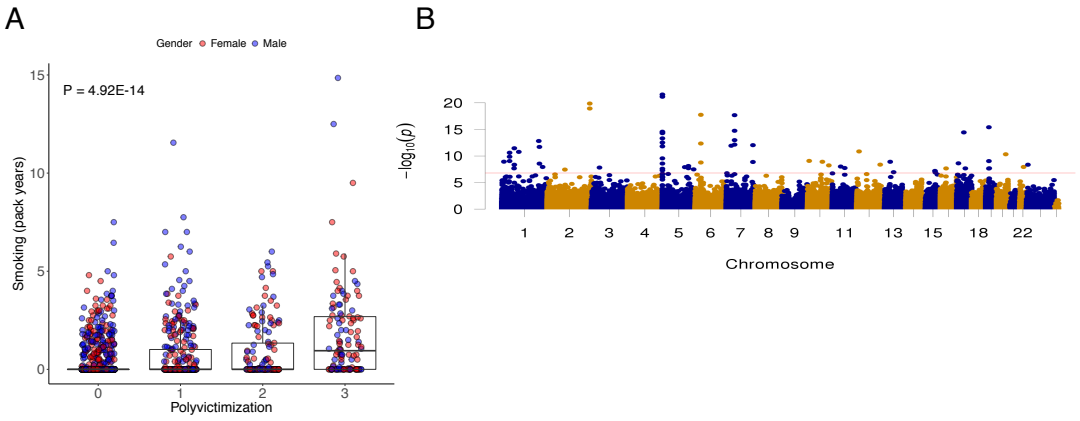


Figure 3-6. Smoking is associated with adolescent polyvictimisation and DNA methylation. **(A)** Smoking pack-years were significantly associated with adolescent polyvictimisation ($P = 4.92 \times 10^{-14}$) in E-Risk, with victimised adolescents more likely to smoke and to smoke more pack-years. **(B)** DNA methylation at 64 probes was significantly associated ($P < 1.61 \times 10^{-7}$) with smoking pack-years in a model regressing DNA methylation onto pack-years, controlling for batch, sex and cell-type composition.

We subsequently performed an EWAS of smoking among E-Risk participants, regressing DNA methylation onto smoking pack-years, while controlling for family-relatedness structure as well as covarying batch, sex and derived blood cell-type proportion. As expected, we identified widespread DNA methylation differences (**Figure 3-6**) associated with smoking exposure, with 64 CpG probes

significant at $P < 1.61 \times 10^{-7}$ (**Table 3-4**). All 64 probes were identified as associated with smoking in large recent EWAS of smoking (Joehanes et al., 2016). Interestingly, there is substantial overlap between the probes identified to be associated with adolescent polyvictimisation and those associated with smoking pack-years in E-Risk, with three of the four array-wide significant probes from the polyvictimisation EWAS – those annotated to *AHRR* (**Table 3-3**) – also found among the 64 probes significantly associated with pack-years.

Table 3-4. 64 probes associated are significantly associated with smoking pack-years at an array-wide significance level ($P < 1.61 \times 10^{-7}$). All 64 probes had previously been reported as significantly associated with smoking by Joehanes et al., 2016.

Probe	Location	Genes	*P* value	Beta (%)
cg21161138	5:399360	*C5orf55; AHRR*	2.75×10^{-22}	-1.27
cg05575921	5:373378	*AHRR; C5orf55*	7.11×10^{-22}	-3.00
cg01940273	2:233284934	*ALPI; ALPPL2*	1.35×10^{-20}	-1.24
cg05951221	2:233284402	*ALPI; ALPPL2*	1.17×10^{-19}	-1.19
cg14753356	6:30720108	*DDR1; IER3*	1.79×10^{-18}	-0.82
cg22132788	7:45002486	*PURB; MYO1G*	2.09×10^{-18}	1.38
cg03636183	19:17000585	*F2RL3*	3.91×10^{-16}	-1.15
cg07826859	7:45020086	*MYO1G*	1.81×10^{-15}	-0.40
cg26703534	5:377358	*C5orf55; AHRR*	2.74×10^{-15}	-1.15
cg19572487	17:38476024	*RARA; GJD3*	3.65×10^{-15}	-0.55
cg14817490	5:392920	*C5orf55; AHRR*	4.15×10^{-15}	-0.83
cg25648203	5:395444	*C5orf55; AHRR*	5.09×10^{-15}	-0.91
cg03991871	5:368447	*AHRR; C5orf55*	4.81×10^{-14}	-1.01
cg12803068	7:45002919	*PURB; MYO1G*	1.05×10^{-13}	1.92
cg11071448	1:202584465	*SYT2; PPP1R12B*	1.52×10^{-13}	-0.59
cg01899089	5:369969	*AHRR; C5orf55*	3.01×10^{-13}	-0.53
cg06126421	6:30720080	*DDR1; IER3*	4.49×10^{-13}	-1.05
cg19089201	7:45002287	*PURB; MYO1G*	7.34×10^{-13}	0.86
cg21322436	7:145812842	*CNTNAP2*	9.23×10^{-13}	-0.41
cg02451831	7:26578098	*SNX10; SKAP2*	1.18×10^{-12}	-0.60
cg23916896	5:368804	*AHRR; C5orf55*	1.54×10^{-12}	-0.77
cg08709672	1:206224334	*AVPR1B*	1.98×10^{-12}	-0.59
cg25189904	1:68299493	*GNG12*	3.51×10^{-12}	-1.18
cg07986378	12:11898284	*BCL2L14; ETV6*	1.39×10^{-11}	-0.71
cg09935388	1:92947588	*GLMN; GFI1*	1.69×10^{-11}	-1.24
cg15542713	1:42385581	*HIVEP3*	2.43×10^{-11}	0.98
cg07339236	20:50312490	*NFATC2; ATP9A*	4.76×10^{-11}	-0.53
cg24049493	1:42385941	*HIVEP3*	1.38×10^{-10}	0.81
cg24688690	5:345850	*AHRR; C5orf55*	2.81×10^{-10}	-0.28

cg04640972	10:8373522	*GATA3*	8.42×10^{-10}	0.45
cg16145216	1:42385662	*HIVEP3*	9.00×10^{-10}	0.70
cg15159987	19:17003890	*F2RL3; CPAMD8*	9.14×10^{-10}	-0.45
cg04885881	1:11123118	*SRM*	1.18×10^{-9}	-0.58
cg03450842	10:80834947	*PPIF; ZMIZ1*	1.23×10^{-9}	-0.49
cg02013841	13:49159767	*CYSLTR2; RCBTB2*	1.23×10^{-9}	-0.63
cg25949550	7:145814306	*CNTNAP2*	1.34×10^{-9}	-0.26
cg15342087	6:30720209	*DDR1; IER3*	1.69×10^{-9}	-0.26
cg05460226	17:8804279	*PIK3R6; PIK3R5*	2.40×10^{-9}	-0.61
cg26529655	5:424371	*C5orf55; AHRR*	2.65×10^{-9}	-0.28
cg20295214	1:206226794	*AVPR1B; C1orf186*	2.75×10^{-9}	-0.42
cg21733098	12:127931219	*TMEM132C*	4.26×10^{-9}	-1.16
cg04224247	X:9984515	*CLCN4; SHROOM2*	4.40×10^{-9}	-0.68
cg26764244	1:68299511	*GNG12*	4.45×10^{-9}	-0.52
cg00950497	10:116393423	*AFAP1L2; ABLIM1*	5.86×10^{-9}	0.55
cg02228160	5:143192067	*NR3C1; YIPF5*	7.65×10^{-9}	0.46
cg08442823	11:45951803	*GYLTL1B; PHF21A*	9.70×10^{-9}	0.60
cg02532700	12:37257404	*NCF4*	1.16×10^{-8}	-0.47
cg21446172	1:223745234	*SUSD4; CAPN8*	1.18×10^{-8}	-0.43
cg08266095	5:123149716	*CSNK1G3; ZNF608*	1.25×10^{-8}	-0.58
cg20505728	3:43800538	*C3orf23; ABHD5*	1.52×10^{-8}	0.31
cg10420527	11:68138505	*PPP6R3; LRP5*	1.83×10^{-8}	-0.32
cg09560590	5:143191663	*NR3C1; YIPF5*	1.98×10^{-8}	0.40
cg23973524	19:18873222	*COMP; CRTC1*	2.06×10^{-8}	0.54
cg19406367	1:66999929	*SGIP1*	2.13×10^{-8}	0.41
cg06395298	17:46651225	*HOXB3*	2.17×10^{-8}	-0.62
cg11824827	16:31075547	*ZNF668*	2.19×10^{-8}	0.50
cg14712058	19:16988083	*F2RL3; SIN3B*	2.28×10^{-8}	-0.39
cg04551776	5:393366	*C5orf55; AHRR*	2.50×10^{-8}	-0.45
cg16624521	5:172064552	*NEURL1B*	3.30×10^{-8}	0.35
cg03604731	2:96824799	*DUSP2; STARD7*	3.63×10^{-8}	-0.22
cg17287155	5:393347	*C5orf55; AHRR*	5.08×10^{-8}	-0.25
cg09651136	15:72525012	*PKM2*	6.45×10^{-8}	0.33
cg23126342	13:67801125	*PCDH9*	1.12×10^{-7}	0.70
cg00310412	15:74724918	*CYP11A1; SEMA7A*	1.26×10^{-8}	-0.30

We next regressed DNA methylation onto adolescent polyvictimisation including smoking pack-years as an additional covariate and found no DMPs passing our stringent array-wide significance threshold (**Figure 3-3**, second panel).

Another prominent challenge to interpreting the association between victimisation and DNA methylation is that the association may be confounded not only by shared family-wide environmental factors, such as poverty, but also by genetic

factors. For example, twin studies reveal that exposure to victimisation is under genetic influence (implying the presence of a gene-environment correlation) (Fisher et al., 2015), and that DNA methylation at a large number of sites across the genome is also heritable (Bell et al., 2011, Zhang et al., 2010, Hannon et al., 2016b, Marzi et al., 2016). Studying exposure-discordant twins is a powerful approach to causal inference in epigenetic epidemiology because it allows researchers to compare same-age siblings who differ in their victimisation exposure but who otherwise grow up in the same family and share family-wide environmental risks (e.g., poverty) as well as twins who are genetically-identical (in the case of MZ twins) (Mill and Heijmans, 2013, Rakyan et al., 2011, Birney et al., 2016). We used a twin-difference design in which we regressed differences in DNA methylation between twins onto differences in victimisation exposure (**Figure 3-7**), controlling for differences in cell-type proportions. We found no array-wide significant associations between adolescent polyvictimisation and DNA methylation in either DZ or MZ twins (**Figure 3-3**, bottom panels).

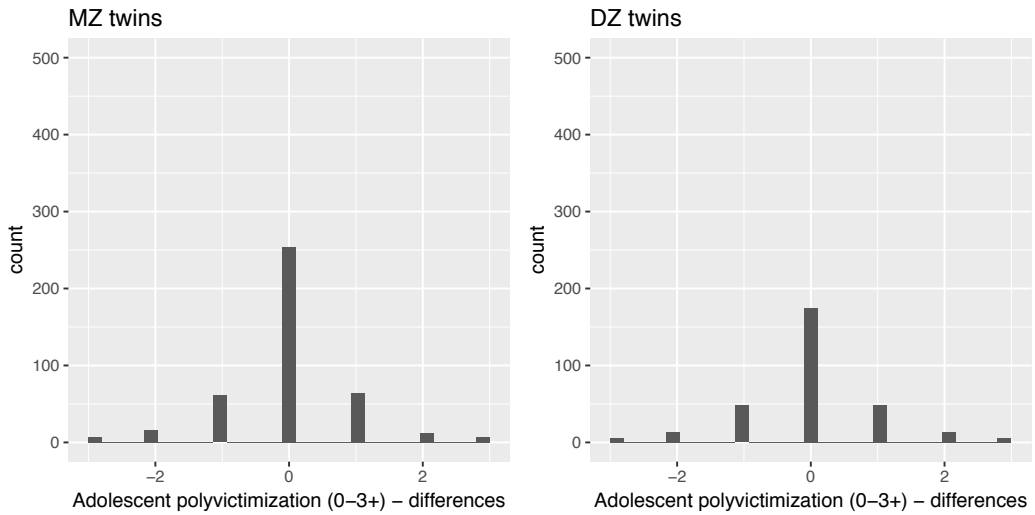


Figure 3-7. Distributions of twin differences in adolescent polyvictimisation.

Of course, not all forms of victimisation are alike. Some types involve physical injury whereas others involve psychological insult; some may be immediate, whereas others remote (e.g., cyberbullying); some are perpetrated by strangers whereas others by people known to the victim. Moreover, some forms of victimisation may have more severe consequences. Measuring polyvictimisation may be diluting the effects of specific forms of victimisation. Thus, we next disaggregated our measure of polyvictimisation into its constituent parts and tested the association between exposure to each victimisation type (i.e. maltreatment, neglect, sexual victimisation, family violence, peer/sibling victimisation, internet/mobile phone victimisation, and crime victimisation; **Table 3-5**) and DNA methylation. The results revealed few novel associations and sparsely distributed significant findings: we detected seven array-wide significant associations across the seven types of victimisation (**Figure 3-8**, **Table 3-6**). Four of these were previously identified in our analysis of adolescent polyvictimisation and only one probe emerged as significant across multiple types of victimisation (cg05575921 annotated to *AHRR* and *C5orf55*). Additional controls for smoking, as discussed above, attenuated these effects (**Figure 3-9, Table 3-6**).

Table 3-5. Distribution of individual victimisation types in adolescence. Shown are the sample distributions across adolescent victimisation types.

Severity	0	1	2	3	4	5	NA
Maltreatment	1435	42	66	60	45	10	0
Neglect	1553	16	15	36	27	10	1
Sexual victimisation	1447	41	58	67	12	31	2
Family violence	1326	15	12	95	129	80	1
Peer/sibling victimisation	686	40	223	447	219	42	1
Internet/mobile phone victimisation	1323	15	100	116	68	35	1
Crime victimisation	803	71	376	83	243	82	0

Table 3-6. Six probes are associated with individual types of victimisation in adolescence, specifically with neglect, peer/sibling victimisation and crime victimisation. None of the associations remain significant when controlling for smoking.

			Not controlled for smoking		Controlled for smoking	
Probe	**Location**	**Genes**	***P***	**Beta (%)**	***P***	**Beta (%)**
Neglect						
cg05446860	20:25129126	*VSX1; ENTPD6*	2.25E-08	-0.39	1.85E-07	-0.38
Peer/sibling victimisation						
cg05575921	5:373378	*AHRR; C5orf55*	2.63E-08	-0.58	1.60E-02	0.21
Crime victimisation						
cg05575921	5:373378	*AHRR; C5orf55*	4.26E-12	-0.78	8.81E-03	-0.24
cg07826859	7:44986611	*MYO1G*	1.02E-09	-0.23	7.61E-06	-0.17
cg26703534	5:430358	*C5orf55; AHRR*	4.78E-09	-0.34	1.23E-02	-0.13
cg21161138	5:452360	*C5orf55; AHRR*	2.41E-08	-0.35	2.87E-02	-0.12
cg06126421	6:30828059	*DDR1; IER3*	1.43E-07	-0.31	3.17E-02	-0.12

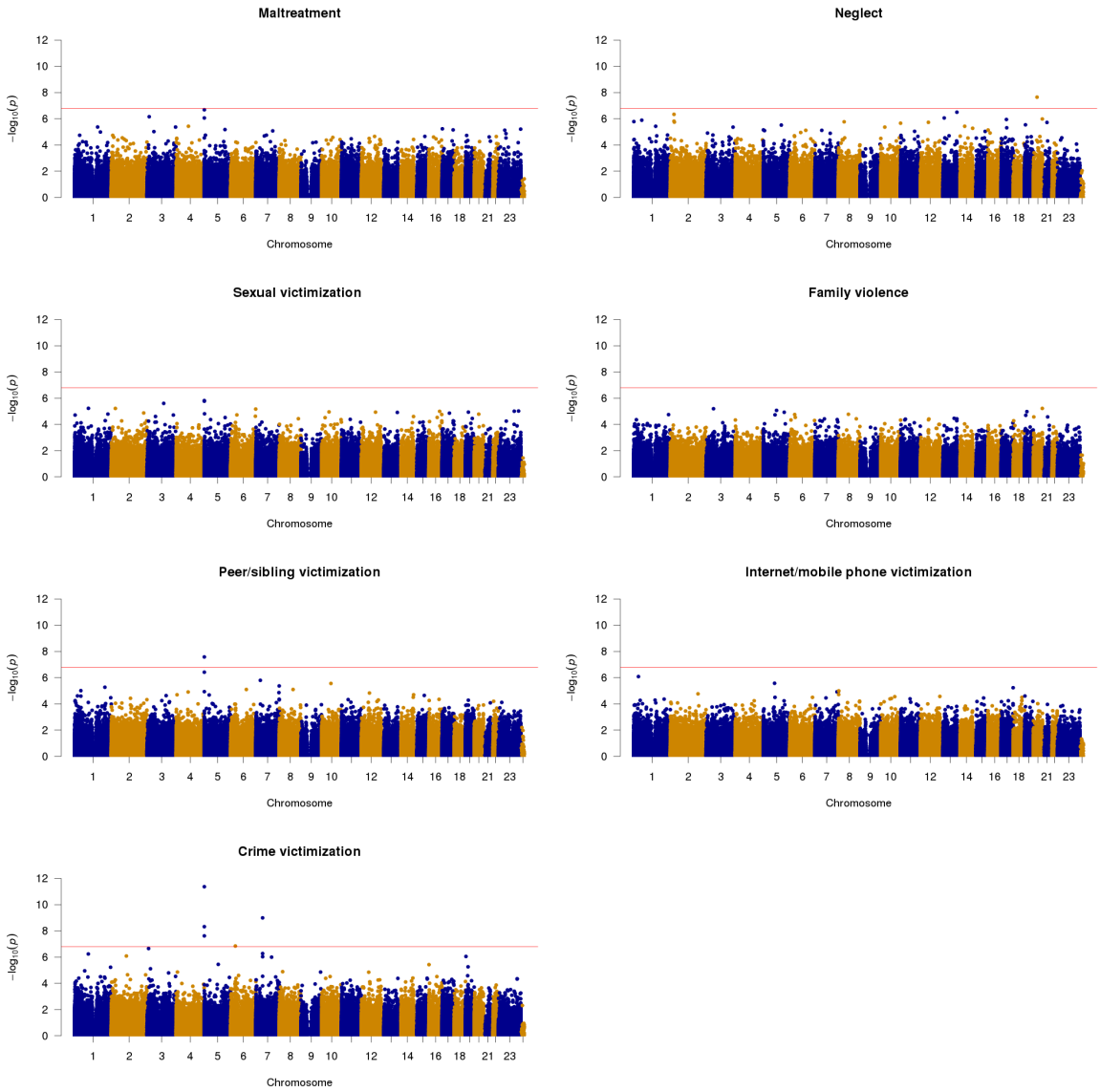


Figure 3-8. Few novel associations and sparsely distributed significant findings were observed for the association between DNA methylation and seven individual types of victimisation. For the seven individual victimisation types constituting the adolescent polyvictimisation measure (i.e., maltreatment, neglect, sexual victimisation, family violence, peer/sibling victimisation, internet/mobile phone victimisation, and crime victimisation) a total of seven array-wide significant associations were observed across three victimisation types (neglect, peer/sibling victimisation and crime victimisation). Four of these associations had been identified in the EWAS of adolescent polyvictimisation (**Figure 3-3**) and DNA methylation at one probe (cg05575921annotated to *AHRR* and *C5orf55*) was significantly associated with multiple types of victimisation.

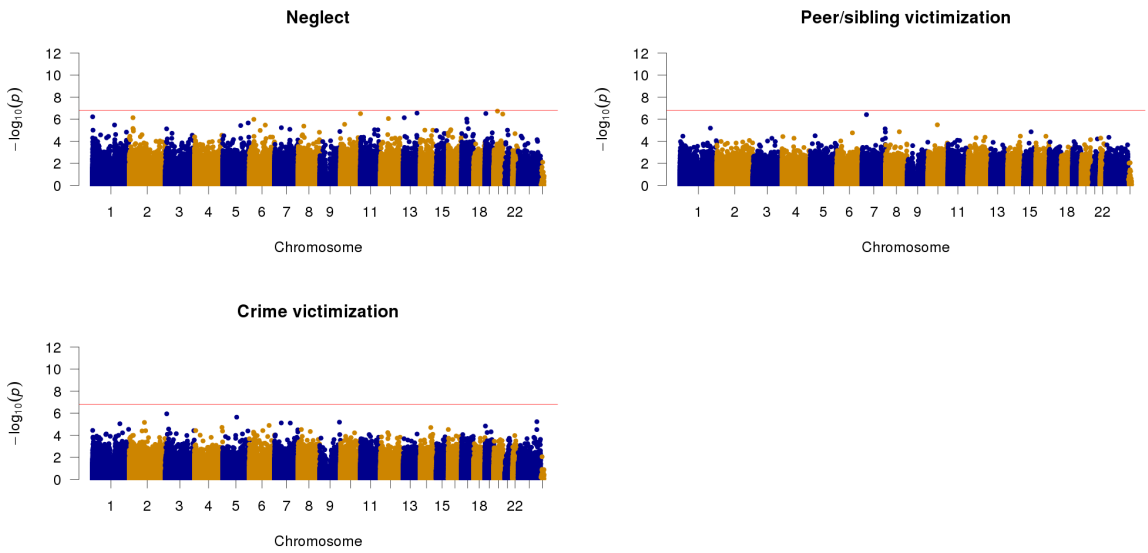


Figure 3-9. Manhattan plots of $-\log_{10}$ *P* values for maltreatment, neglect and crime victimisation in adolescence, controlling for smoking pack-years. No probes pass the array-wide significance threshold ($P < 1.61 \times 10^{-7}$) in any of the three analyses.

3.3.2 Testing the sensitive period of childhood victimisation

Although victimisation experiences peak in adolescence, it is hypothesised that the most biologically consequential victimisation is experienced earlier in life (Andersen et al., 2008). To test for such a sensitive period to victimisation, we next tested whether children victimised in the first decade of life exhibited differences in DNA methylation that were detectable in blood at age 18, regressing DNA methylation onto childhood polyvictimisation accounting for family structure and including batch, sex and cell-type proportion estimates as covariates.

DNA methylation at one CpG probe (cg10318825 annotated to *TMEM120A* and *POR*) was associated with childhood polyvictimisation at an array-wide significance level ($P = 1.02 \times 10^{-7}$, **Figure 3-10,** top panel). The association between polyvictimisation and cg10318825 was attenuated when controlling for smoking, although another probe (cg10911283 annotated to *FTMT*) reached array-wide significance ($P = 1.38 \times 10^{-7}$, **Figure 3-10**, second panel). Of note, effect sizes for the top 100 ranked DMPs identified in the model not controlling for smoking are highly correlated with the effect sizes controlling for smoking ($r = 0.966$, $P < 2.2 \times 10^{-16}$, **Figure 3-11**). No probes passed the array-wide

significance threshold in the DZ or MZ twin difference models (**Figure 3-12**; **Figure 3-10**, bottom panels).

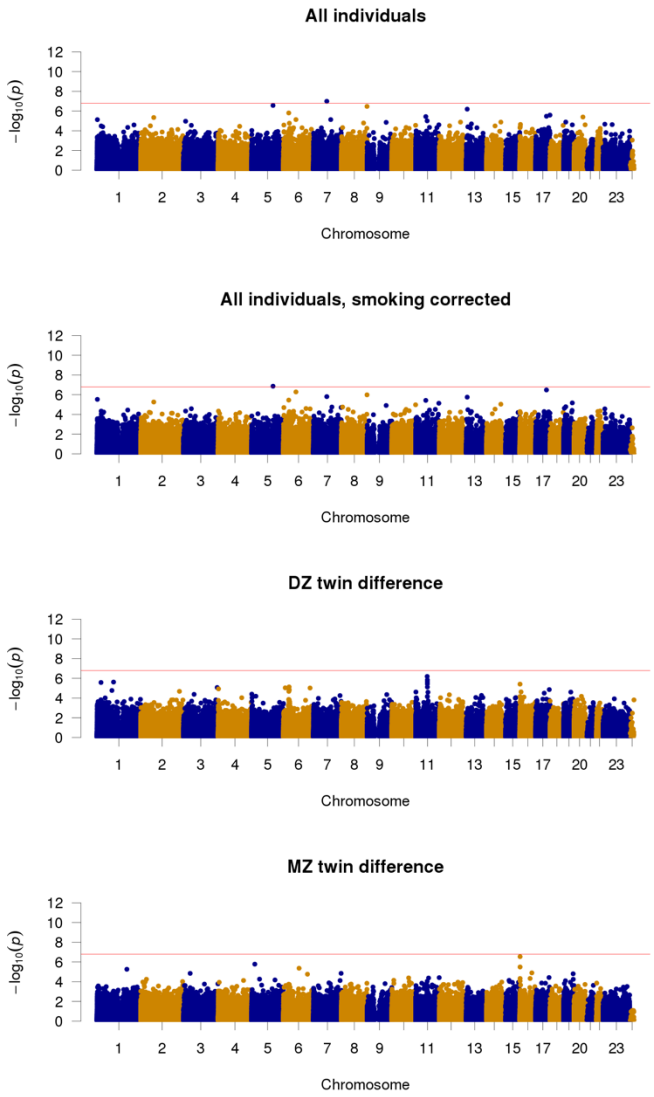


Figure 3-10. No consistent associations between DNA methylation and childhood polyvictimisation are observed across multiple models. The four panels show $-\log_{10} P$ values for the associations between DNA methylation and childhood polyvictimisation. **Top panel:** Only one probe is array-wide significant (cg10318825 annotated to *TMEM120A* and *POR*, $P = 1.02 \times 10^{-7}$) in a regression model controlling for batch, sex and cell-type proportion estimates. **Second panel:** A different probe passes the array-wide significance threshold when adding smoking pack-years as a further covariate (cg10911283 annotated to *FTMT*, $P = 1.38 \times 10^{-7}$). The bottom panels show *P* value distributions from two twin difference models, which examine the association between twin-pair differences in DNA methylation and differences in polyvictimisation, while controlling for differences in cell-type proportion estimates. No significant associations were observed for either the DZ (**third panel**) or MZ twin difference model (**fourth panel**).

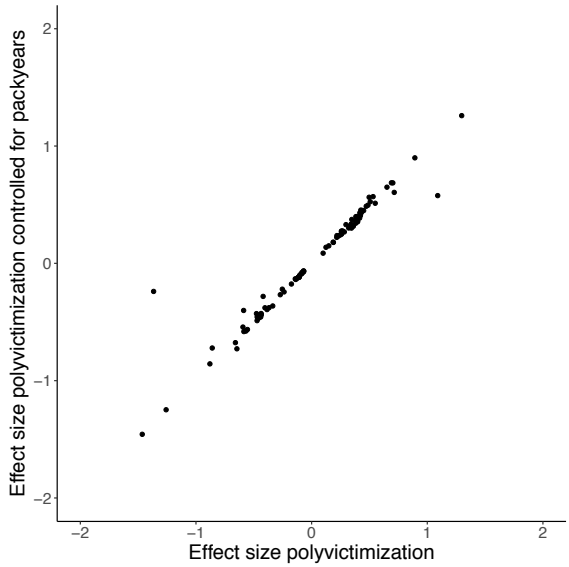


Figure 3-11. Effect sizes of the top 100 probes identified in the regression of DNA methylation on childhood polyvictimisation not controlling for smoking. Shown are the effect sizes in the model not controlling for smoking (x-axis) compared to those found in the model including smoking pack-years as a covariate (y-axis).

Figure 3-12. Distributions of twin differences in childhood polyvictimisation.

Table 3-7. Distribution of individual victimisation types in childhood. Shown are the sample distributions across childhood victimisation types.

Severity	0	1	2	NA
Physical abuse	1319	253	86	0
Physical neglect	1506	123	29	0
Emotional neglect	1465	140	53	0
Sexual victimisation	1629	15	14	0
Domestic violence	871	493	294	0
Peer/sibling victimisation	918	582	156	2

As with adolescent polyvictimisation, we disaggregated our measure of childhood polyvictimisation into its constituent parts and tested the association between exposure to each victimisation type (i.e. physical abuse, physical neglect, emotional neglect, sexual victimisation, domestic violence and peer/sibling victimisation; **Table 3-7**) and DNA methylation. A total of nine probes passed the array-wide threshold for childhood sexual victimisation (**Table 3-8**), while none of the other childhood victimisation types were significantly associated with DNA methylation (**Figure 3-13**). Seven of the nine associations between DNA methylation and childhood sexual victimisation remained significant when controlling for smoking (cg00713005 annotated to *TSSK3*, cg06927775 annotated to *ATP6V0A1*, cg04577451 annotated to *KCTD5* and *PRSS27*, cg20974621 annotated to *HHATL* and *HIGD1A*, cg06454157 annotated to *CCR6* and *FGFR1OP*, cg08755156 annotated to *RHEB* and *PRKAG2*, and cg17898054 annotated to *SLC2A1* and *EBNA1BP2*) and an additional three probes became significant in the model controlling for smoking (cg11087774 annotated to *NKX6-2* and *TTC40*, cg19897017 annotated to *MAD1L1* and *ELFN1*, and cg01954930 annotated to *DHFRL1*; **Figure 3-14, Table 3-8**). These findings may indicate that sexual victimisation is associated with stable DNA methylation differences in whole blood in young adulthood, beyond what is detectable with the aggregated polyvictimisation measure. However, these findings should be interpreted with caution because only a small group of children were sexually victimised (n = 29).

Table 3-8. Nine probes are significantly associated with childhood sexual victimisation. Seven of these associations remain significant when controlling for smoking and three additional associations become significant in the smoking-controlled model.

			Childhood sexual victimisation		Controlled for smoking	
Probe	**Location**	**Genes**	***P***	**Beta (%)**	***P***	**Beta (%)**
cg00713005	1:32824095	*TSSK3*	5.07E-09	0.59	6.45E-09	0.61
cg06927792	19:54975886	*LENG9*	8.17E-09	0.29	1.46E-08	0.29
cg04577497	12:10766264	*MAGOHB*	1.08E-08	0.32	3.35E-09	0.34
cg20974659	11:1375304	*TOLLIP; BRSK2*	1.58E-08	2.14	3.16E-08	2.11
cg06454226	18:2572965	*METTL4; NDC80*	4.83E-08	-1.12	1.09E-07	-1.14
cg09856107	3:12701776	*RAF1; MKRN2*	8.15E-08	1.40	2.49E-07	1.38
cg08755218	16:67261029	*TMEM208*	1.02E-07	-0.26	8.12E-08	-0.26
cg11525620	19:35417684	*ZNF30*	1.20E-07	0.59	1.28E-06	0.56
cg17898069	1:234445489	*C1orf31; SLC35F3*	1.43E-07	-1.08	1.07E-07	-1.10
cg11087819	8:142838482	*FLJ43860; TSNARE1*	1.81E-07	1.50	1.25E-07	1.51
cg19897071	5:35991382	*UGT3A1*	3.99E-07	-2.47	3.79E-08	-2.64
cg01954957	14:38263617	*TTC6*	4.78E-07	1.35	5.84E-08	1.41

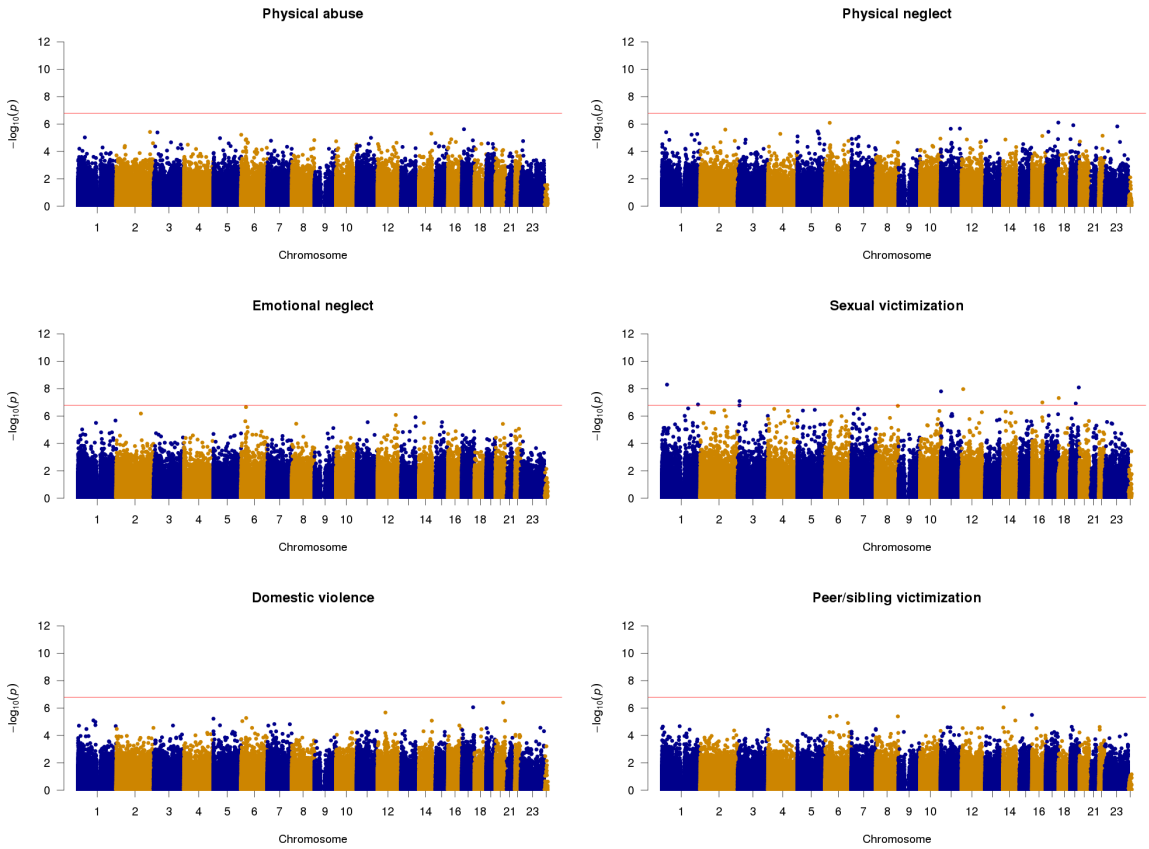


Figure 3-13. Sexual victimisation is the only victimisation type in childhood which was found to be significantly associated with DNA methylation. For the six individual victimisation types constituting the childhood polyvictimisation measure (i.e. physical abuse, physical neglect, emotional neglect, sexual victimisation, domestic violence and peer/sibling victimisation) only sexual victimisation showed any array-wide significant associations with DNA methylation. A total of nine probes passed the array-wide threshold for childhood sexual victimisation (**Table 3-9**).

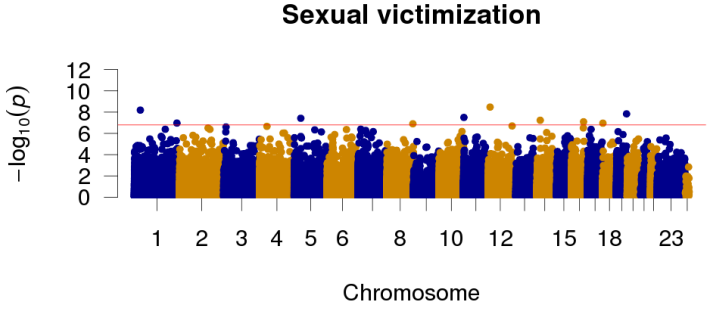


Figure 3-14. Manhattan plots of $-\log_{10}$ *P* values for sexual victimisation in childhood, controlling for smoking pack-years.

3.3.3 Testing the cumulative stress load hypothesis

Perhaps the most consequential stressors are those that are experienced chronically or that reoccur. A striking finding in epidemiology is the phenomenon of revictimisation (Widom et al., 2008). For example, in E-Risk, every type of victimisation in childhood was associated with a significantly greater risk of victimisation in adolescence, and polyvictimised children were 60% more likely to be polyvictimised in adolescence (RR = 1.60, (95% CI: 1.42, 1.82)). This raises the possibility that the biological embedding of victimisation is especially like to occur in response to a greater cumulative stress load.

Using our longitudinal data about victimisation over the first two decades of life - across childhood and through adolescence - a latent class analysis (see **Methods**) identified four types of individuals:

1) The first group consisted of individuals who were not exposed to childhood or adolescent victimisation (group 0, n = 834 in the EWAS sample). 2) Childhood domestic violence defined the victimisation profile of n = 308 individuals (group 1, n = 254 in the EWAS sample). In this group, 89% of individuals experienced moderate to severe domestic violence as a child; 3) Peer victimisation defined the violence profile of n = 529 individuals (group 2, n = 412 in the EWAS sample). 88% of individuals classified into this group experienced moderate or severe peer victimisation during adolescence; 4) Exposure to multiple moderate or severe types of violence defined the victimisation profile of n = 206 individuals (group 3, n = 158 in the EWAS sample). 36% of individuals in this group suffered two or more types of severe victimisation during childhood and 68% experienced two or more types of severe victimisation during adolescence (**Table 3-1**). We tested the cumulative stress load hypothesis by treating these four victimisation categories as a factor, and testing the for differences in DNA methylation between each victimisation group (groups 1-3) compared to the non-victimised individuals (group 0). No probes showed differential methylation between groups 1 or 2 and the non-victimised individuals, while six probes passed the array-wide significance threshold when comparing the most severely victimised individuals (group 3) with non-victimised individuals ($P < 1.61 \times 10^{-7}$, **Figure 3-15**, third panel). These included the four probes previously found to be associated with adolescent polyvictimisation (**Table 3-1, Table 3-9**). None of these six probes remained array-wide significant when controlling for smoking, but another probe

(cg13854030 annotated to *SPIN1* and *CDK20*) reached array-wide significance in this analysis. Sensitivity analyses performed with a different scaling of cumulative victimisation revealed similar results (**Table 3-9**).

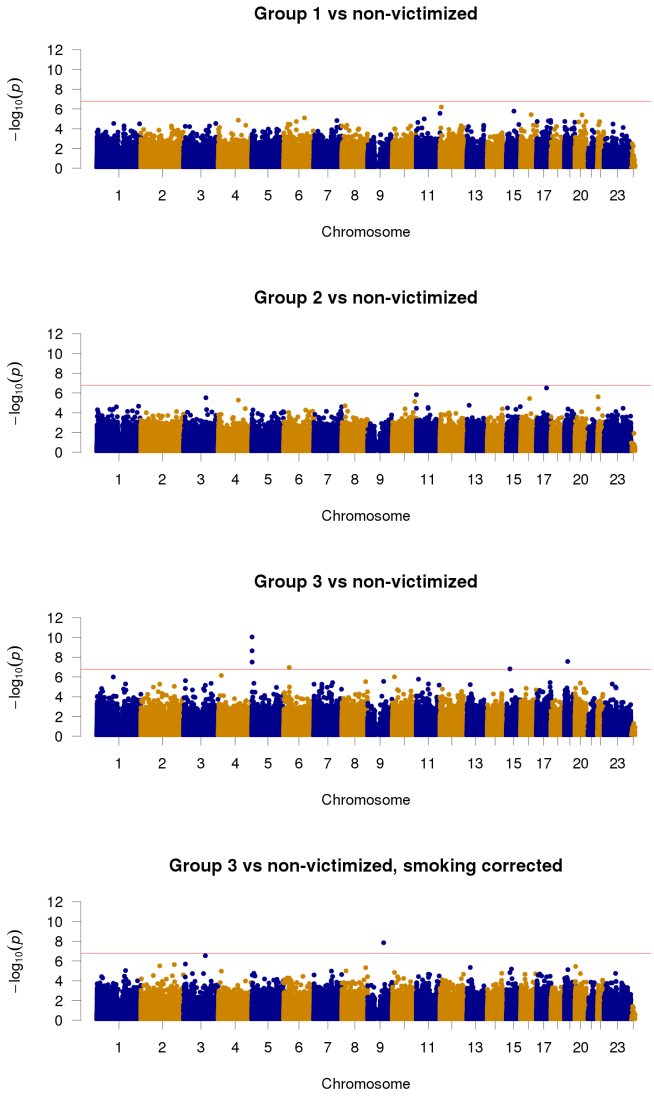


Figure 3-15. The association between DNA methylation and cumulative polyvictimisation is confounded by smoking. The four panels show $-\log_{10} P$ values for the associations between DNA methylation and cumulative polyvictimisation. No probes pass the array-wide multiple testing threshold ($P < 1.61 \times 10^{-7}$) in a regression model controlling for batch, sex and cell-type proportion estimates when comparing groups 1 (**top panel**) or 2 (**second panel**) with the non-victimised group. **Third panel**: Six probes are array-wide significant when comparing the most severely victimised individuals (group 3) with non-victimised individuals. These include the four probes previously identified to be significantly associated with adolescent polyvictimisation (**Figure 3-3**). **Bottom panel**: None of the six probes identified as significantly differentially methylated between group 3 and the non-victimised subgroup remain significant when additionally controlling for smoking. One additional probe passes the array-wide multiple testing threshold in this analysis.

Table 3-9. Eight probes are significantly associated with two different cumulative victimisation measures. Of these, three probes (cg05575921, cg26703534 and cg21161138) are significant across both models.

Probe	Location	Genes	*P*	Beta (%)
LCA score –group 3 vs non-victimised individuals				
cg05575921	5:373378	*AHRR; C5orf55*	8.84E-11	-5.35
cg26703534	5:430358	*C5orf55; AHRR*	2.22E-09	-2.37
cg03636183	19:17000585	*F2RL3*	2.67E-08	-2.67
cg21161138	5:452360	*C5orf55; AHRR*	3.06E-08	-2.44
cg06126421	6:30828059	*DDR1; IER3*	1.08E-07	-2.22
cg00944304	15:40738072	*CHST14, BAHD1*	1.51E-07	-1.16
Additive model				
cg05575921	5:373378	*AHRR; C5orf55*	7.71E-12	-1.14
cg21161138	5:452360	*C5orf55; AHRR*	1.95E-09	-0.54
cg26703534	5:430358	*C5orf55; AHRR*	2.90E-09	-0.50
cg10318825	7:75611554	*TMEM120A; POR*	2.92E-09	0.11
cg14179389	1:92947961	*GLMN; GFI1*	9.81E-09	-0.71

3.3.4 Epigenetic interrogation of stress-related genes

A hypothesis-free approach to the study of stress overlooks cumulative evidence about biological plausibility of candidate genes. In particular, genes involved in HPA-axis reactivity may be differentially methylated in response to victimisation. Genes widely hypothesised to play a role in mediating stress-reactivity include *NR3C1* (the glucocorticoid receptor, which is primarily involved in response to inflammation (McGowan et al., 2009)); *FKBP5* (an important regulator of the glucocorticoid receptor network (Klengel et al., 2013)); *BDNF* (the brain derived neurotrophic factor, a member of the nerve growth factor family (Roth et al., 2009, Unternaehrer et al., 2012)); *AVP* (the gene encoding the neuropeptide vasopressin (Murgatroyd et al., 2009), which is secreted in as part of the HPA response to stress (Meyer-Lindenberg et al., 2011)), and *CRHR1* (the corticotrophin releasing hormone receptor, another major player in the HPA pathway (Todkar et al., 2015, Polanczyk et al., 2009, Bradley et al., 2008)). In addition, genetic variants in *SLC6A4* (the serotonin transporter gene), have been implicated in stress reactivity and some evidence suggests that these effects may be mediated by altered DNA methylation (Alexander et al., 2014, Duman and Canli, 2015, Provenzi et al., 2016). Evidence from severely deprived children has

also flagged *CYP2E1* (Kumsta et al., 2016), a member of the cytochrome P450 super family of enzymes, with a role in the metabolism of various exogenous compounds including drugs of abuse and neurotoxins (Strobel et al., 2001).

We identified all probes on the Illumina 450K array annotated to each of these seven genes (see **Methods**). The number of probes per gene ranged from 25 to 140. Within each gene, we tested the association between levels of DNA methylation and victimisation in childhood and adolescence, controlling for sex, batch and cell-type proportion estimates. While none of the probes annotated to any of these seven genes reached array-wide significance in any of the genome-wide analyses reported here, we recognise the importance of also reporting individual gene-level results for these candidate genes.

To illustrate, **Figure 3-16** shows a plot of $\log_{-10}$ *P* values and effect directions (negative [hypo-] vs. positive [hyper-] methylation) for all probes annotated to *NR3C1*, with two thresholds: the array-wide significance level and a gene-wide Bonferroni-corrected significance level (0.05/number probes annotated to the gene and included in this analysis). For *NR3C1*, we observed no gene-wide significant probes associated with either adolescent polyvictimisation or childhood polyvictimisation. The regional Manhattan plots of the other genes are shown in **Figure 3-17-Figure 3-22** and individual association statistics for all probes annotated to these seven genes are reported in **Appendix A**. Overall, only one probe reached gene-wide significance across all further six genes and two victimisation periods: cg00140191, a probe upstream of *FKBP5*, passed gene-wide significance for the association with childhood polyvictimisation ($P = 7.63 \times 10^{-4}$). Together, these data are not supportive of robust changes in DNA methylation in the vicinity of previously nominated candidate genes in victimised young people.

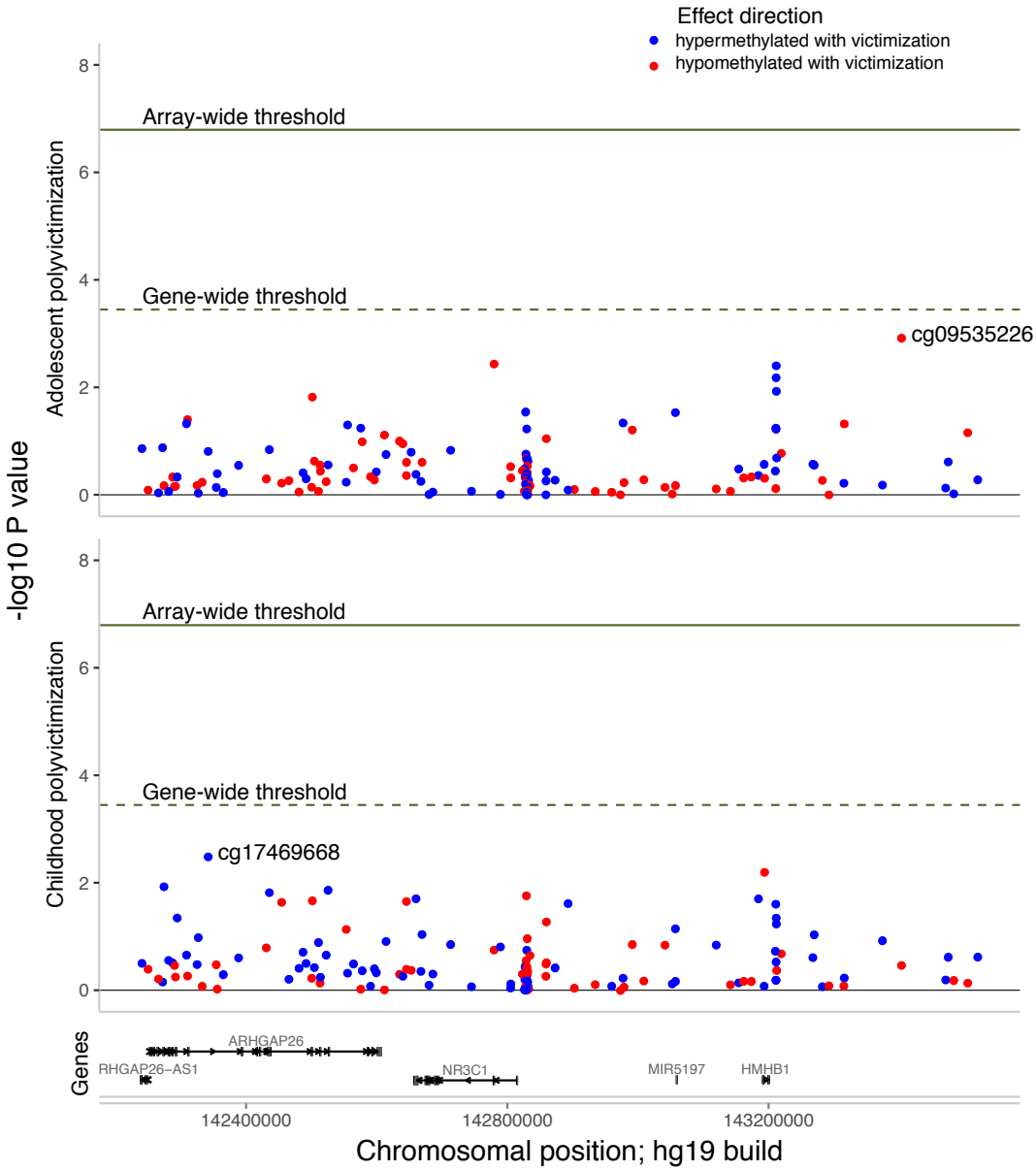


Figure 3-16. Polyvictimisation in childhood or adolescence is not associated with differential methylation across probes annotated to *NR3C1*. Regional Manhattan plot of -$\log_{10}$ *P* values for all 140 probes annotated to *NR3C1*, showing summary statistics from the associations between DNA methylation and adolescent polyvictimisation (**top panel**) and childhood polyvictimisation (**bottom panel**). Effect directions are indicated by blue (hypermethylated) and red (hypomethylated) colour. Of the 140 probes annotated to *NR3C1* none pass the array- or gene-wide thresholds for either of the two victimisation exposures.

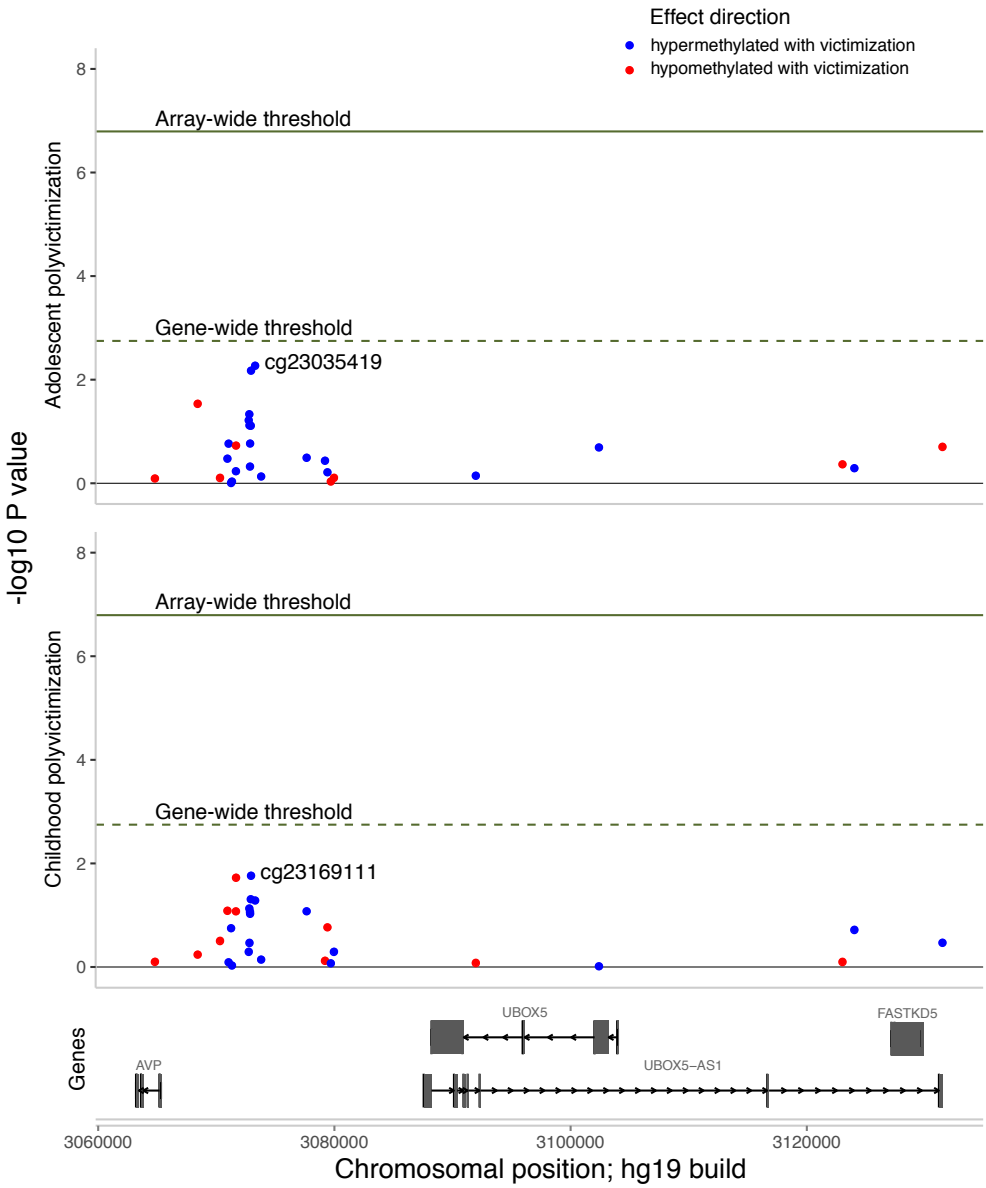


Figure 3-17. Polyvictimisation in childhood or adolescence is not associated with differential methylation across probes annotated to *AVP*. Regional Manhattan plot of -$\log_{10}$ *P* values for all probes annotated to *AVP*, showing summary statistics from the associations between DNA methylation and adolescent polyvictimisation (**top panel**) and childhood polyvictimisation (**bottom panel**). Effect directions are indicated by blue (hypermethylated) and red (hypomethylated) colour. Of the 28 probes annotated to *AVP* none pass the array- or gene-wide thresholds for either of the two victimisation exposures.

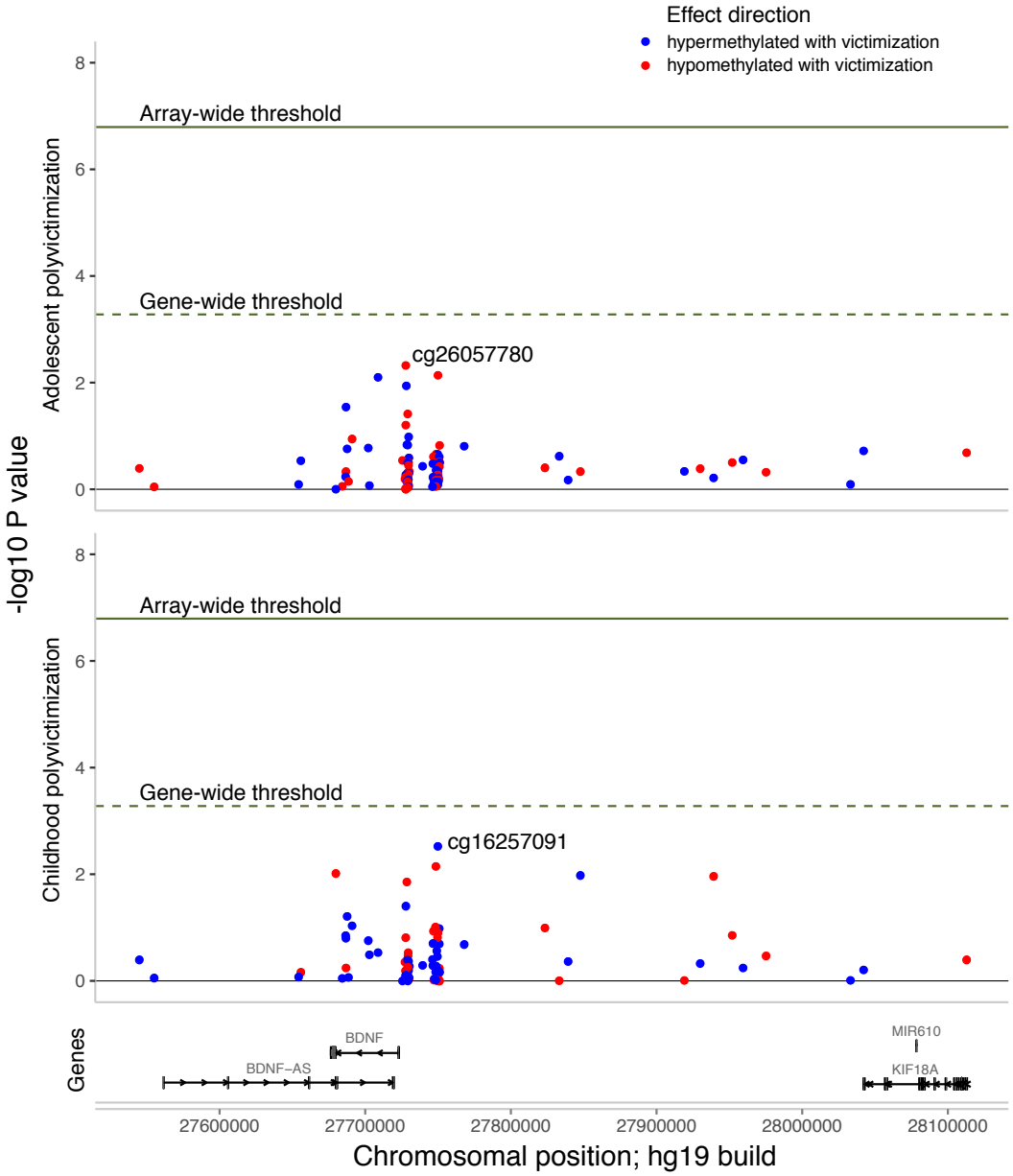


Figure 3-18. Polyvictimisation in childhood or adolescence is not associated with differential methylation across probes annotated to *BDNF*. Regional Manhattan plot of -$\log_{10}$ *P* values for all probes annotated to *BDNF*, showing summary statistics from the associations between DNA methylation and adolescent polyvictimisation (**top panel**), childhood polyvictimisation (**bottom panel**). Effect directions are indicated by blue (hypermethylated) and red (hypomethylated) colour. Of the 95 probes annotated to *BDNF* none pass the array- or gene-wide thresholds for either of the two victimisation exposures.

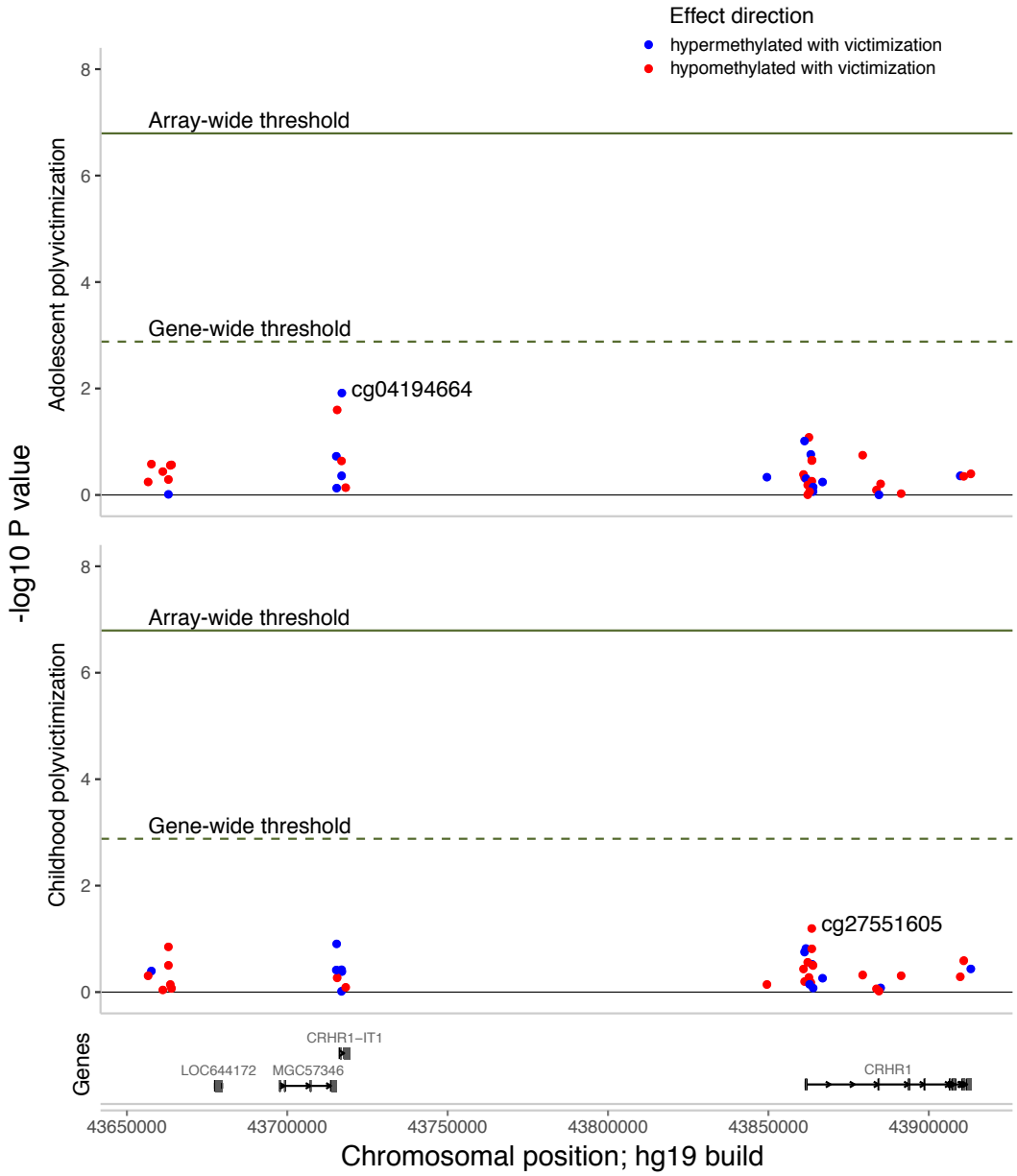


Figure 3-19. Polyvictimisation in childhood or adolescence is not associated with differential methylation across probes annotated to *CRHR1*. Regional Manhattan plot of -$\log_{10}$ *P* values for all probes annotated to *CRHR1*, showing summary statistics from the associations between DNA methylation and adolescent polyvictimisation (**top panel)** and childhood polyvictimisation (**bottom panel**). Effect directions are indicated by blue (hypermethylated) and red (hypomethylated) colour. Of the 38 probes annotated to *CRHR1* none pass the array- or gene-wide thresholds for either of the two victimisation exposures.

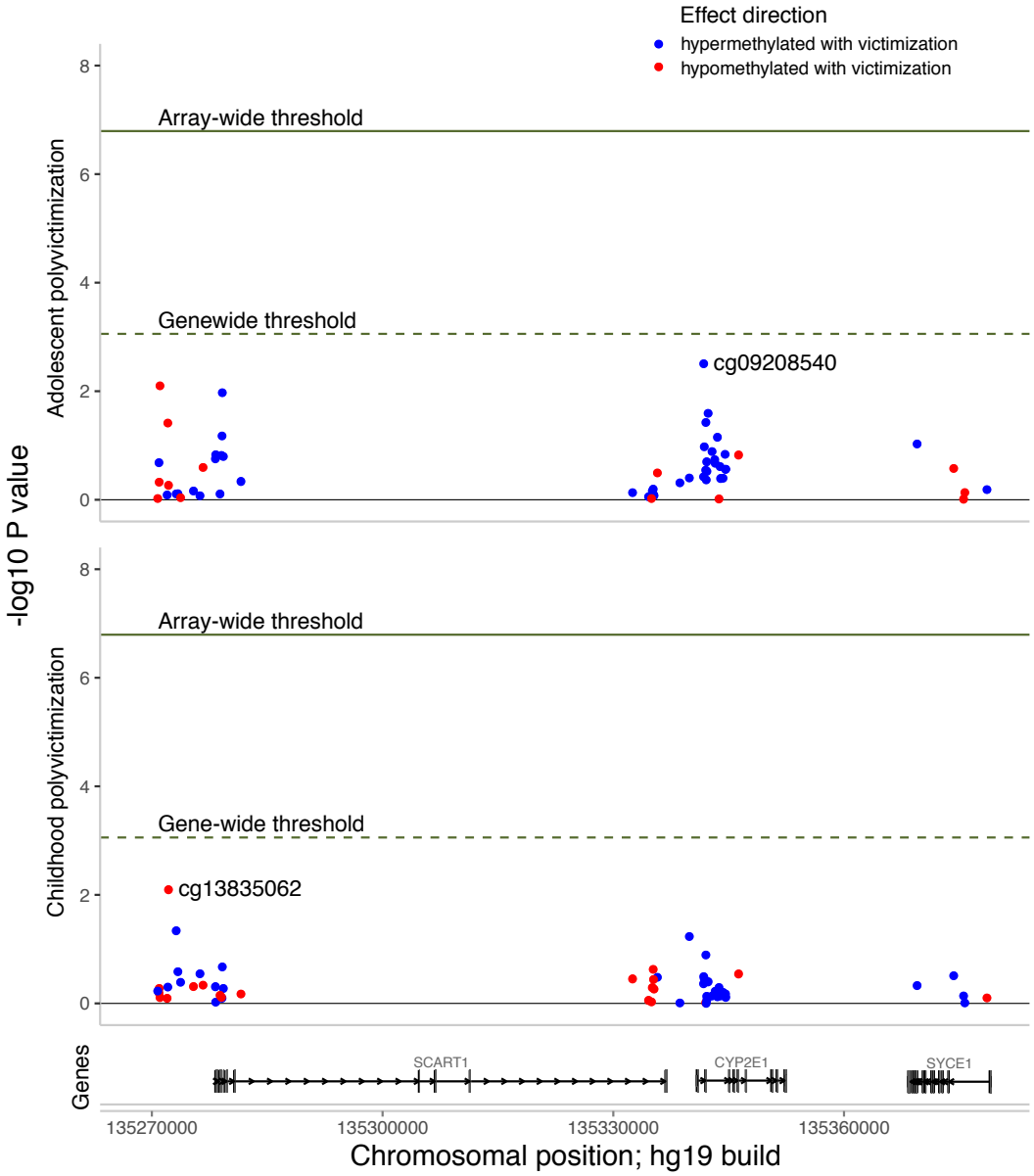


Figure 3-20. Polyvictimisation in childhood or adolescence is not associated with differential methylation across probes annotated to *CYP2E1*. Regional Manhattan plot of -$\log_{10}$ *P* values for the all probes annotated to *CYP2E1*, showing summary statistics from the associations between DNA methylation and adolescent polyvictimisation (**top panel**) and childhood polyvictimisation (**bottom panel**). Effect directions are indicated by blue (hypermethylated) and red (hypomethylated) colour. Of the 57 probes annotated to *CYP2E1* none pass the array- or gene-wide thresholds for either of the two victimisation exposures.

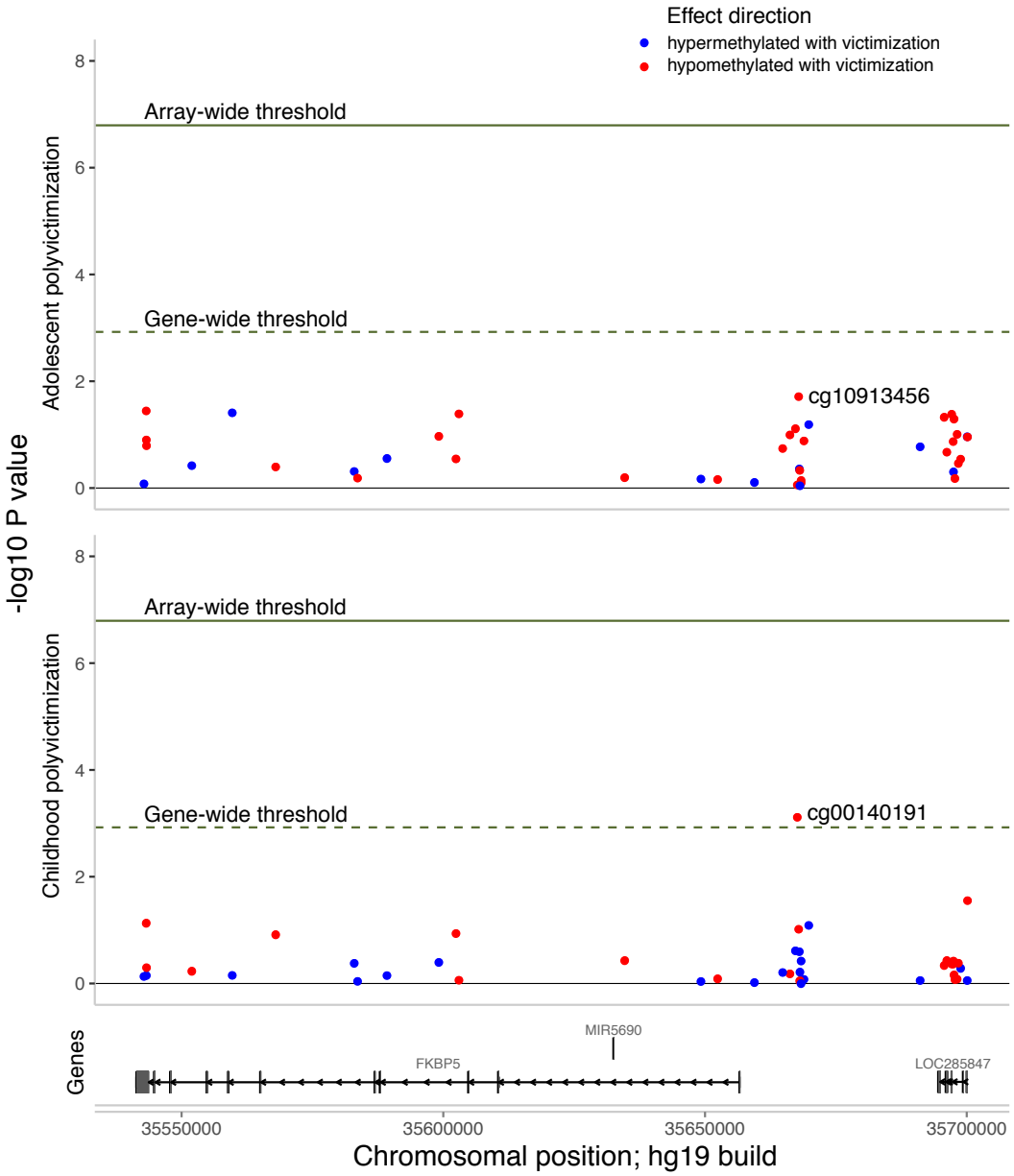


Figure 3-21. Polyvictimisation in adolescence is not associated with differential methylation across probes annotated to *FKBP5*. Regional Manhattan plot of -$\log_{10}$ *P* values for all 42 probes annotated to *FKBP5*, showing summary statistics from the associations between DNA methylation and adolescent polyvictimisation (**top panel**) and childhood polyvictimisation (**bottom panel**). Effect directions are indicated by blue (hypermethylated) and red (hypomethylated) colour. Of the 42 probes annotated to *FKBP5* none pass the array-wide threshold for adolescent polyvictimisation. One probe, cg00140191, passes the gene-wide threshold for childhood polyvictimisation, exhibiting hypomethylation with victimisation exposure.

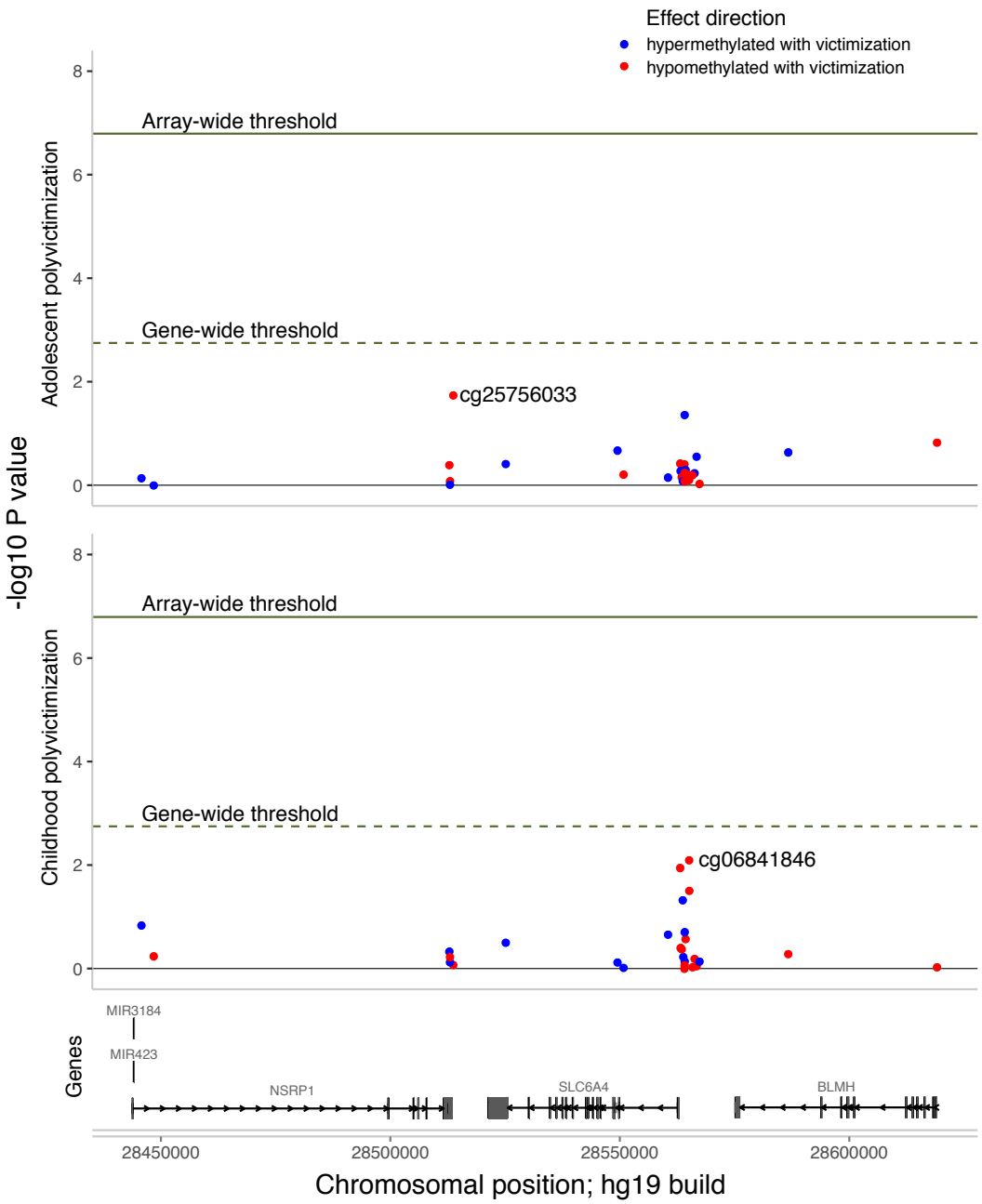


Figure 3-22. Polyvictimisation in adolescence or childhood is not associated with differential methylation across probes annotated to *SLC6A4*. Regional Manhattan plot of -$\log_{10}$ *P* values for the all probes annotated to *SLC6A4*, showing summary statistics from the associations between DNA methylation and adolescent polyvictimisation (**top panel**) and childhood polyvictimisation (**bottom panel**). Effect directions are indicated by blue (hypermethylated) and red (hypomethylated) colour. Of the 28 probes annotated to *SLC6A4* none pass the array-wide threshold for either of the two victimisation exposures.

3.3.5 Does victimisation accelerate epigenetic ageing?

Stress is hypothesised to accelerate epigenetic ageing. A promising biomarker of aging is the "epigenetic clock", a highly accurate multi-tissue predictor of chronological age based on DNA methylation patterns at 353 CpG sites (Horvath, 2013). Individuals whose DNA methylation age exceeds their chronological age are thought to be ageing faster, and faster epigenetic ageing has been shown to predict all-cause mortality (Marioni et al., 2015). More recently, stressful life events and stress-related phenotypes have also been linked to accelerated epigenetic ageing (Zannas et al., 2015, Boks et al., 2015, Wolf et al., 2016).

We tested whether polyvictimisation is associated with accelerated epigenetic ageing by regressing DNA methylation age acceleration (the difference between DNA methylation age and chronological age) onto polyvictimisation. DNA methylation age acceleration was not significantly associated with polyvictimisation in adolescence (P = 0.42), childhood (P = 0.24) or cumulatively (minimum P of the three victimisation groups = 0.38; **Figure 3-23**).

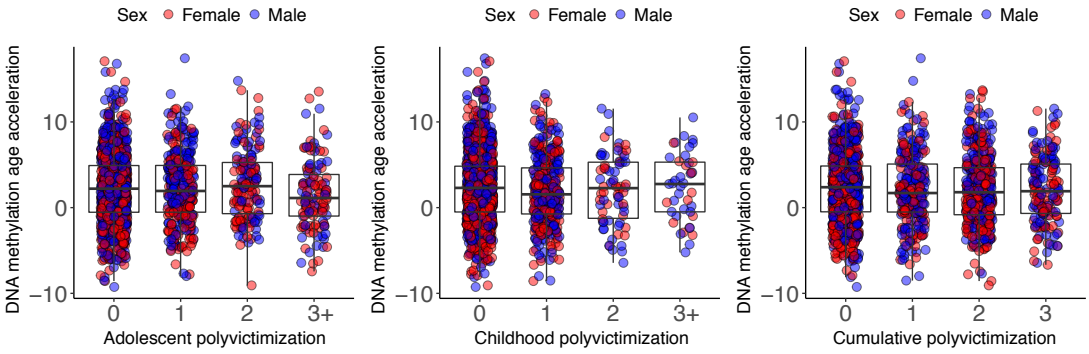


Figure 3-23. Distributions of DNA methylation age acceleration by polyvictimisation in three different exposure time periods. DNA methylation age acceleration is not significantly associated with polyvictimisation in adolescence (P = 0.42), childhood (P = 0.24) or cumulatively (minimum P = 0.38).

3.4 Discussion

This study represents, to our knowledge, the most comprehensive analysis yet undertaken in humans of epigenetic variation associated with victimisation experiences across the first two decades of life. Using a large longitudinal population-based cohort of twins, we investigated associations between victimisation across childhood and adolescence and DNA methylation isolated from whole blood at age 18. Overall, we find little evidence to support widespread differences in DNA methylation associated with early-life exposure to stress. The results of our study have important implications for analyses of epigenetic variation associated with psycho-social exposures in population-based cohorts. In particular, our study highlights the importance of controlling for potential confounders, especially exposure to tobacco smoke, in such analyses.

First, while we find statistically significant associations ($P < 1.61 \times 10^{-7}$) between DNA methylation at four Illumina 450K probes (cg05575921, cg26703534, cg21161138 and cg14179389) and adolescent as well as cumulative victimisation, we show that these associations are confounded by smoking exposure. There are no statistically significant DNA methylation differences between DZ or MZ twin pairs discordant for victimisation in adolescence or childhood. Second, across different exposure periods only one probe, cg10911283, shows a significant association between childhood polyvictimisation and DNA methylation when controlling for smoking, but DNA methylation at this probe is not significantly associated with childhood polyvictimisation in the model not controlling for smoking or in the DZ or MZ twin difference models. Third, with the exception of childhood sexual victimisation, no significant associations between individual types of victimisation in adolescence or childhood and DNA methylation are found, which survive smoking-correction. Childhood sexual victimisation is significantly associated with DNA methylomic variation at nine probes, seven of which remain significant when correcting for smoking. However, this association is to be interpreted with caution because of the small sample size (only n = 29 exposed individuals). A follow-up project based on a sample with greater numbers of exposed individuals could shed more light onto this association.

Many current studies of epigenetic variation associated with early-life stress in human populations have focussed on specific *a priori* candidate genes – for

example *NR3C1* (McGowan et al., 2009, Tyrka et al., 2012, Martin-Blanco et al., 2014, Guillemin et al., 2014), *FKBP5* (Non et al., 2016, Yehuda et al., 2016, Tyrka et al., 2015) and *SLC6A4* (Beach et al., 2010, Ouellet-Morin et al., 2013, Booij et al., 2015). We therefore undertook a detailed investigation of candidate stress genes, finding no robust evidence for altered DNA methylation associated with victimisation exposure in probes annotated to any of these candidate loci. Only one probe out of a total 428 tested pass our relatively lenient gene-wide thresholds: cg00140191 annotated to *FKBP5* for childhood polyvictimisation (P = 7.63 10^{-4}) and it would not have been significant at a Bonferroni threshold correcting for all 428 probes ($P < 1.17 \times 10^{-4}$). Notably, a number of probes would be classed as "significant" at the nominal P value threshold often employed for such candidate gene analyses, highlighting that many of these previous studies are likely to be reporting false positive associations. Furthermore, while some recent publications have linked stressful life events and trauma, to an increase in DNA methylation age (Zannas et al., 2015, Wolf et al., 2016, Boks et al., 2015), we were not able to replicate these findings for victimisation exposure and accelerated epigenetic ageing.

Taken together, we find very little evidence for robust associations between early life victimisation exposure and DNA methylation in whole blood at age 18, beyond those associations confounded by smoking. Even the sparse individual array-wide significant probes need to be viewed with caution, as our analyses involved many (possibly correlated) analyses and there is an increased possibility of false positives.

Our study has several specific strengths and advantages that improve its validity and strengthen the evidence supporting our conclusions: The E-Risk Longitudinal Twin Study is one of the most detailed and thorough studies of child and adolescent development and the consequences of early life adverse exposures (Fisher et al., 2015). By oversampling high-risk families E-Risk achieves a population-representative sample of early-life development, underlining its external validity. To our knowledge this study represents the largest study of the epigenetic consequences early-life stress yet undertaken and we are well powered to detect variation in DNA methylation associated with external exposures. This is exemplified by the fact that we detect 64 array-wide significant probes associated with smoking pack-years. Furthermore, we employed an

embedded twin difference model on large subsamples of our full sample (307 DZ twin pairs and 425 MZ twin pairs), allowing us to control for confounding caused by the family environment as well as genetic influences. Previous studies of epigenetic variation in humans associated with early-life stress, in contrast, have been relatively small (Guillemin et al., 2014, Romens et al., 2015, McGowan et al., 2009, Labonte et al., 2012, Yehuda et al., 2016, Booij et al., 2015, Suderman et al., 2012); given their predominant focus on candidate genes in a limited number of samples, it is likely that many of the previously reported associations in the literature represent false-positive findings.

One of the most important questions concerning our results therefore is how to interpret them in the greater context of the existing literature. This question is particularly salient given existing reports that document DNA methylation differences in peripheral tissues following stress. First, with studies of mental health and psychiatric phenotypes the organ of main interest is usually the brain. In our longitudinal study, we are limited to whole blood as a surrogate tissue. While it is a heterogeneous tissue, there are several validated reference-based estimators (Houseman et al., 2012, Horvath, 2013) controlling for cell-type composition. In addition, the stress response of the HPA-axis has a direct impact on the immune response and T cell composition in blood (Palacios and Sugawara, 1982, Elenkov, 2004) and it would therefore be plausible to see differences in DNA methylation associated with severe stress exposure, such as the one caused by victimisation experiences. However, we did not find any robust associations between victimisation and DNA methylation in whole blood and it is possible that blood does not represent a useful surrogate tissue for this research question.

Second, there may be issues relating to our stress exposure measures leading us to observe different findings to previous studies. The measures of victimisation exposure reported for E-Risk are thorough and detailed, spanning different exposure periods across the first two decades of life. We disaggregate the polyvictimisation measures in childhood and adolescence into its constituent parts at a comprehensive severity coding level, making it possible to uncover effects of individual victimisation types that might have been diluted by the composite measure. Additionally, we consider multiple scalings of the cumulative victimisation variable (LCA model, additive model; **Table 3-2**) thereby minimising

the effect any particular operationalisation of this exposure could have on the outcome. The victimisation measures in E-Risk are robustly associated with adverse physical and mental health outcomes (Arseneault et al., 2006, Fisher et al., 2012, Jaffee et al., 2004, Baldwin et al., 2016, Fisher et al., 2015), substantiating the validity of our variables. DNA methylation data was collected and processed through a streamlined and stringent pipeline. These data are of high quality as evidenced by a filtering rate of < 1% of samples in our QC pipeline, and our confirmation of known smoking-associated probes (Joehanes et al., 2016).

Third, by using a population representative sample we allow maximum external validity of our study. In addition, we use a longitudinal study design with minimal attrition, so exposures are unlikely to be associated with study-related confounding factors. This comes at the cost of lower numbers of exposed individuals in some of the victimisation categories. Sexual victimisation in childhood is characterised by particularly low numbers of exposed individuals and the significant associations between DNA methylation and sexual victimisation should be interpreted with caution in consequence. However, for our main polyvictimisation variables, we have substantial numbers of exposed individuals, with 593 individuals having experienced at least one type of victimisation in adolescence and 118 of those exposed to three or more different types of victimisation. We also looked at victimisation across different time periods, disaggregated our polyvictimisation measures into individual victimisation exposures and used a literature-informed candidate gene approach with gene-wide *P* value thresholds. However, despite the variety of angles from which we studied the association between early-life victimisation and DNA methylation, we find no robust evidence for an association.

Fourth, while E-Risk is a longitudinal study and our exposure and mental health measures are collected over time, whole blood was collected at a single time point (age 18) and the DNA methylation data is therefore cross-sectional. Unfortunately, blood samples from earlier periods of life are not available for this study, but the longitudinal collection of DNA methylation profiles, to evaluate changes in DNA methylation in relation to exposures is a promising future avenue of research. We have little understanding of the temporal effects of stress and methylation, and it is possible that the lag in our research, looking at effects of

victimisation which took place between age 5 and 18, is simply not the correct temporal scale.

Fifth, by nature of the technology we are limited in our study of DNA methylation to the probes featured on the Illumina 450K HumanMethylation Array. This is a subset of all CpG sites in the genome and the relevant sites may not be featured on this microarray. More generally, while epigenetic modifications may play a mediating role in translating and establishing the long-term mental health outcomes associated with early-life stress and victimisation, DNA methylation may not be a modification directly involved in this process.

Sixth, there may be problems with the previously reported epigenetic associations with early-life stress. Much of the previous work was based on animal models and investigated DNA methylation changes in the brain (Weaver et al., 2004, Kundakovic et al., 2013, Desarnaud et al., 2008, Kember et al., 2012). These may not directly translate to human epidemiology, because of differences in biology or experimental design, but could also include false positives as sample sizes in these animal studies tend to be extremely small, often with five or fewer animals per exposure group (Weaver et al., 2004, Desarnaud et al., 2008, Daniels et al., 2009, Henningsen et al., 2012, Doherty et al., 2016). Similarly, with few exceptions in recent years (Mehta et al., 2013, van der Knaap et al., 2014, Martin-Blanco et al., 2014, Beach et al., 2010), work in human epigenetic epidemiology has often used small sample sizes (Guillemin et al., 2014, Naumova et al., 2012, Yehuda et al., 2016, Kumsta et al., 2016, Borghol et al., 2012). This is particularly true for studies performed on brain tissue (McGowan et al., 2009, Labonte et al., 2012, Suderman et al., 2012). To complicate matters further, some studies were performed on clinical populations with limited generalisability (Steiger et al., 2013, Melas et al., 2013, Martin-Blanco et al., 2014). The majority of studies, including nearly all animal work, used suboptimal targeted technologies including targeted bisulfite sequencing or bisulfite pyrosequencing which only allow the investigation of very small regions of methylation and increase the likelihood of false positives in the context of publication bias (Weaver et al., 2004, Desarnaud et al., 2008, Daniels et al., 2009, Kember et al., 2012, Liberman et al., 2012, Kundakovic et al., 2013, Kosten et al., 2014, Steiger et al., 2013, Tyrka et al., 2012, Melas et al., 2013, van der Knaap et al., 2014, Martin-Blanco et al., 2014, Romens et al., 2015, Guillemin et

al., 2014, McGowan et al., 2009, Labonte et al., 2012, Non et al., 2016, Ouellet-Morin et al., 2013, Doherty et al., 2016, Saunderson et al., 2016, Beach et al., 2010, Beach et al., 2011, Murgatroyd et al., 2009, Roth et al., 2009, Kinnally et al., 2010, Booij et al., 2015, Yehuda et al., 2016). Even studies using array technologies sometimes employed forms of targeted or variance based filtering to overcome sub-threshold results (Esposito et al., 2016, Mehta et al., 2013). In addition, interpretation of significance levels in a greater genomic context is often prevented by only reporting *P* value thresholds (e.g. $P < 0.05$) rather than the raw *P* values themselves, often observed in combination with an omission of multiple-testing correction (Weaver et al., 2004, Desarnaud et al., 2008, Kember et al., 2012, Kundakovic et al., 2013, Tyrka et al., 2012, Martin-Blanco et al., 2014, McGowan et al., 2009, Labonte et al., 2012, Non et al., 2016, Naumova et al., 2012, Doherty et al., 2016, Saunderson et al., 2016, Murgatroyd et al., 2009, Borghol et al., 2012, Suderman et al., 2012, Yehuda et al., 2016). Some studies used very lenient thresholds for FDR correction (Esposito et al., 2016, Suderman et al., 2012) or applied more lenient candidate gene thresholds post hoc (Weder et al., 2014). Interestingly, three previous studies using array-wide profiling did not report differential methylation in any of the HPA-axis candidate stress genes investigated here (*AVP*, *BDNF*, *CRHR1*, *FKBP5*, *NR3C1*) at array-wide multiple testing thresholds (Houtepen et al., 2016, Weder et al., 2014, Kumsta et al., 2016).

In summary, even in the context of the existing literature, we find that there is insufficient evidence to conclude that there is a robust association between early-life victimisation and DNA methylation detectable in whole blood. With regard to genomic and epigenomic studies, our results demonstrate the importance to adequately address for confounders, including smoking, cell-type heterogeneity, familial environments and genetics. Our findings also highlight the fact that genome- or array-wide methods in the affected tissue are warranted to understand the neurobiology of stress and its long-term mental health impact. Finally, perhaps we need to come to terms with the possibility that experimental work and animal models are not necessarily well matched to human epigenetic epidemiology in uncovering the molecular pathways linking stress exposures to long-term adverse sequelae.

4. Tissue-specific patterns of allelically-skewed DNA methylation

4.1 Publication

EPIGENETICS
2016, VOL. 11, NO. 1, 24–35
http://dx.doi.org/10.1080/15592294.2015.1127479

Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

OPEN ACCESS

Tissue-specific patterns of allelically-skewed DNA methylation

Sarah J. Marzi[a], Emma L. Meaburn[b], Emma L. Dempster[c], Katie Lunnon[c], Jose L. Paya-Cano[a], Rebecca G. Smith[c], Manuela Volta[a], Claire Troakes[a], Leonard C. Schalkwyk[d], and Jonathan Mill[a,c]

[a]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; [b]Department of Psychological Sciences, Birkbeck, University of London, London, UK; [c]University of Exeter Medical School, University of Exeter, Exeter, UK; [d]School of Biological Sciences, University of Essex, Colchester, UK

ABSTRACT

While DNA methylation is usually thought to be symmetrical across both alleles, there are some notable exceptions. Genomic imprinting and X chromosome inactivation are two well-studied sources of allele-specific methylation (ASM), but recent research has indicated a more complex pattern in which genotypic variation can be associated with allelically-skewed DNA methylation in *cis*. Given the known heterogeneity of DNA methylation across tissues and cell types we explored inter- and intra-individual variation in ASM across several regions of the human brain and whole blood from multiple individuals. Consistent with previous studies, we find widespread ASM with > 4% of the ~220,000 loci interrogated showing evidence of allelically-skewed DNA methylation. We identify ASM flanking known imprinted regions, and show that ASM sites are enriched in DNase I hypersensitivity sites and often located in an extended genomic context of intermediate DNA methylation. We also detect examples of genotype-driven ASM, some of which are tissue-specific. These findings contribute to our understanding of the nature of differential DNA methylation across tissues and have important implications for genetic studies of complex disease. As a resource to the community, ASM patterns across each of the tissues studied are available in a searchable online database: http://epigenetics.essex.ac.uk/ASMBrainBlood.

ARTICLE HISTORY
Received 26 August 2015
Revised 19 November 2015
Accepted 25 November 2015

KEYWORDS
Allele-specific DNA methylation; blood; brain; cerebellum; cortex; epigenetics; genomic imprinting; SNP

Introduction

DNA methylation is the most widely studied and stable epigenetic mark across the mammalian genome, playing a key role in the developmental regulation of gene expression. DNA methylation is generally symmetrical across both alleles, although exceptions characterized by allelic asymmetry include differentially methylated regions (DMRs) regulating the monoallelic expression of genes associated with X chromosome inactivation in females and genomic imprinting.[1-5] Recently, it has been shown that the allelic-skewing of DNA methylation can also be driven by DNA sequence variation, with methylation quantitative trait loci (meQTLs) predominantly acting in *cis*.[6-11] ASM can be regarded as a special case of intermediate DNA methylation (IM), which has been found to occur in regions spanning a large portion of the human genome. It has been estimated that ASM contributes up to 18% of IM in the human genome.[12]

DNA methylation patterns are highly dynamic during normal development and cellular differentiation[13-16] and tissue-specific patterns of DNA methylation have been widely studied in humans.[17-20] In complex tissues such as the brain, for example, DNA methylation differentiates between functionally distinct regions[21,22] and cell types.[16,23-26] Patterns of IM can also be tissue-specific,[12] with growing evidence for the widespread prevalence of tissue-specific ASM.[27,28] In mouse, for example, it has been reported that 28% of imprinted genes are monoallelically expressed in a single tissue type, often the brain or extra-embryonic tissue.[29] Examples of tissue-specifically imprinted genes include *KCNQ1*, which becomes biallelically expressed in embryonic heart development,[30] *GNAS*, which is maternally expressed in a wide-range of tissues including the anterior pituitary, thyroid and ovaries but biallelically expressed in others, such as bone and visceral adipose tissue,[31,32] and *GRB10*, which is maternally expressed in most peripheral tissues but paternally expressed in the brain.[29,33] Genetic influences on DNA methylation can also be tissue-specific, with meQTLs determining allelic patterns of methylation in *cis* in certain tissues or cell types.[11,34]

Increasing evidence supports a role for inter-individual variation in DNA methylation in the etiology and pathogenesis associated with a diverse range of complex disease phenotypes.[35] Allelic differences in DNA methylation may be particularly important in this regard, acting as endophenotypes of genetic variation or additional epi-allelic layers mediating the functional consequences of genotypic variation.[36,37] Teasing apart genetic and non-genetic effects in a tissue- and cell type-specific manner will be a crucial step in understanding the association between non-coding genetic variation, DNA methylation, and complex disease.

To investigate the role of tissue-specific variation of ASM in the human brain and its relation to allelic biases in whole blood, we examined ASM across multiple brain regions and matched blood samples collected from multiple donors. Our data shows

CONTACT Jonathan Mill J.Mill@exeter.ac.uk
Supplemental data for this article can be accessed on the publishers website.

that although a large proportion of ASM is conserved across tissues, there are specific differences in the extent and distribution of ASM sites between regions of the brain and whole blood. Genome browser tracks displaying ASM signals as well as an online tool plotting ASM for sites of interest are available for download from a searchable database (http://epigenetics.essex.ac.uk/ASMBrainBlood).

Results

DNA methylation is allelically-skewed at specific locations across the genome

The majority of the genome is not characterized by notable allelic biases in DNA methylation in any of the tissues assessed in this study. The array-wide average of ASM score (in absolute values) is consistently low (mean = 0.025, range = 0.023 to 0.030) (Fig. S1A, Fig. S1B and Table S1). As expected, there is, however, evidence for allelically-biased DNA methylation at a notable number of specific genomic regions; in total 9,311 (4.22%) of the 220,449 informative SNPs in our assay show evidence for allelic-skewing of DNA methylation, defined by an absolute ASM score ≥ 0.10 in at least one tissue and individual. The percentage of amplicons characterized by an ASM score ≥ 0.10 in each of the 21 profiled samples is given in Table S2. The top-ranked loci showing evidence for allelically-skewed DNA methylation in whole blood, cerebellum, and cortex (BA9), and cerebellum are listed in Tables 1–3. Genome Browser tracks and an online ASM database are available from our laboratory website (http://epigenetics.essex.ac.uk/ASMBrainBlood).

Patterns of ASM in whole blood overlap with those identified in a previous study

In a previous study we characterized allelically-skewed DNA methylation in whole blood derived from 5 monozygotic twin pairs.[7] There is a highly significant correlation between absolute ASM scores across all probes informative in both data sets (n = 129,559, $r = 0.21$, $P < 1.0 \times 10^{-50}$, Fig. S2), even though the majority of the probes do not exhibit ASM. Of the 2,704 ASM loci identified in Schalkwyk et al., 1,717 (63.50%) are informative in the current study, with a highly significant cross-study correlation of ASM scores at these probes ($r = 0.52$, $P < 1.0 \times 10^{-50}$). Likewise, there is a highly significant correlation between ASM scores across the two studies at sites showing allelically-skewed DNA methylation in blood in the current study which were also informative in our previous study (r = 0.38, P = 3.0 × 10-28). Of the 15 top-ranked blood ASM sites identified in our current study (Table 1), 7 of the 9 sites (78%) also informative in our previous study of ASM in blood[7] were characterized by an absolute ASM score ≥ 0.10 in both analyses. These data confirm the validity of the MSNP approach for identifying allelically-skewed DNA methylation, reinforcing our previous conclusions about the extent of ASM in whole blood.[7]

The extent and distribution of ASM differs across tissues

The average proportion of informative sites characterized by allelically-skewed DNA methylation (absolute ASM score ≥ 0.10) in each of the 8 tissues profiled was examined (Fig. 1A). Table 4 lists the top-ranked consistently allelically-skewed probes across cerebellum, whole blood, and cortex (BA9), with specific examples shown in Fig. 2A and Fig. 2B. Allelically-skewed DNA methylation appears to be consistently less prevalent in cortical regions (informative probes with ASM score $\geq 0.10 = 0.54\%$) compared to the cerebellum (1.14%) and whole blood (0.84%). The elevated level of allelically-skewed DNA methylation in the cerebellum and whole blood relative to cortex is more pronounced at more extreme ASM score thresholds (i.e., ASM score ≥ 0.20, cortex = 0.003%, cerebellum = 0.019%, whole blood = 0.013%) (Fig. S1C, Fig. S1D and Fig. S3). Of note, there is little variation in the prevalence and distribution of ASM scores between different regions of the cortex (average correlation between 2 cortical areas = 0.52, Fig. 1B and Fig. S4). We therefore selected one representative cortical region (BA9) for inclusion in subsequent analyses. In contrast, we find more striking differences between cortex, cerebellum and whole blood samples with inter-tissue correlations ranging from $r = 0.42$ to 0.48 (Figs. 1C-E). Table 5 lists the probes showing the highest level of variation in ASM scores across tissues with specific

Table 1. Top 15 ASM sites in whole blood, ASM score averaged across individuals.

Rank	SNP ID	Location	Associated gene(s)	Schalkwyk et al. (2010)	Blood ASM score	Cerebellum ASM score	BA9 ASM score
1	SNP_A-2180729 (rs10276966)	7p15.2	*HIBADH, EVX1*[b]	0.29	0.25	0.01	0.03
2	SNP_A-8438077 (rs585451)	15q21.2	*ATP8B4, DTWD1*	NA	0.25	0.01	0.03
3	SNP_A-1841543 (rs10234308)	7p15.3	*MGC87042*	0.34	0.23	0.25	0.23
4	SNP_A-1946136 (rs927000)	20q13.12	*STK4*	NA	0.23	0.04	0.01
5	SNP_A-8450837 (rs4404067)	16q12.2	*SLC6A2, LPCAT2*	0.28	0.22	0.05	0.04
6	SNP_A-4279002 (rs335554)	1q41	*CENPF, KCNK2*	NA	0.22	0.08	0.09
7	SNP_A-8450539 (rs10916799)	1p36.12	*CAMK2N1, LOC339505*	0.18	0.22	0.11	0.15
8	SNP_A-8633222 (rs13165930)	5q33.3	*CCNJL*	NA	0.22	0.02	0.12
9	SNP_A-8472169 (rs11193683)	10q23.1	*NRG3*	0.01	0.22	0.09	0.07
10	SNP_A-2071005 (rs4687210)	3q28	*UTS2D*	NA	0.22	0.16	0.17
11	SNP_A-2185394 (rs852454)	7p22.1	*RNF216*	0.06	0.22	0.11	0.11
12	SNP_A-8702215 (rs12978286)	19p13.11	*FCHO1, MAP1S*	0.19	0.22	0.00	0.07
13	SNP_A-1807006 (rs10481354)	8p23.3	*CLN8, DLGAP2*[a]	0.22	0.22	0.00	0.05
14	SNP_A-8326632 (rs1542180)	2q31.1	*HOXD3, HOXD4*	NA	0.21	0.24	0.19
15	SNP_A-2008150 (rs869108)	11p15.1	*OTOG, MYOD1, USH1C*	0.22	0.21	0.16	0.07

[a] Known imprinted gene
[b] Suspected imprinted gene

Table 2. Top 15 ASM sites in cerebellum, ASM score averaged across individuals.

Rank	SNP ID	Location	Associated gene(s)	Cerebellum ASM score	Blood ASM score	BA9 ASM score
1	SNP_A-4255628 (rs959246)	18q12.3	*SLC14A2, SETBP1*	0.30	0.03	0.05
2	SNP_A-8696273 (rs1003533)	5q31.1	*C5orf56*	0.29	0.16	0.07
3	SNP_A-1820553 (rs7959070)	12q22	*CLLU1OS, BTG1*	0.26	0.03	0.05
4	SNP_A-1841543 (rs10234308)	7p15.3	*MGC87042*	0.25	0.23	0.23
5	SNP_A-2002282 (rs12246813)	10q22.1	*COL13A1, C10orf35*	0.24	0.10	0.10
6	SNP_A-8625237 (rs12493005)	3q26.32	*TBL1XR1*	0.24	0.20	0.24
7	SNP_A-8326632 (rs1542180)	2q31.1	*HOXD3, HOXD4*	0.24	0.21	0.19
8	SNP_A-2160121 (rs7205794)	16q23.3	*CDH13, MPHOSPH6*	0.24	0.01	0.01
9	SNP_A-2052542 (rs3098382)	5q13.2	*MAP1B*	0.24	0.04	0.19
10	SNP_A-2040586 (rs17097827)	14q32.2	*BCL11B, C14orf177*	0.23	0.10	0.18
11	SNP_A-8420373 (rs10186346)	2p22.1	*TMEM178*	0.23	0.01	0.07
12	SNP_A-1788157 (rs7158663)	14q32.2	*MEG3*[a]	0.23	0.01	0.03
13	SNP_A-4235630 (rs9641549)	7q31.2	*TFEC, MDFIC*	0.23	0.05	0.13
14	SNP_A-4264458 (rs1358229)	4q27	*TRPC3, KIAA1109*	0.23	0.01	0.00
15	SNP_A-1931666 (rs6707698)	2q34	*IKZF2, ERBB4*	0.22	0.03	0.02

[a] Known imprinted gene

examples shown in Fig. 3A and Fig. 3B. We used clonal bisulfite sequencing to validate tissue-specific ASM identified by the MSNP method in these 2 regions (Fig. 3C and Fig. 3D), confirming the patterns observed in our array data for both loci.

Informative MSNP probes within DNase I hypersensitive regions are characterized by elevated ASM scores

Enrichment analyses were performed using a Kruskal-Wallis rank-sum test for ASM rank differences between the annotated genic regions (see **Materials and Methods**). We observed a differential distribution of ASM scores across annotated genic regions (i.e., coding, 5′UTR, intergenic, intron, promoter, 3′UTR) in cortex (BA9) ($P = 1.29 \times 10^{-15}$), cerebellum ($P = 3.98 \times 10^{-14}$), whole blood ($P = 2.06 \times 10^{-12}$), and the cross-tissue analysis ($P = 2.12 \times 10^{-26}$). Post-hoc tests identified these differences to be primarily driven by an enrichment of high ASM scores in promoter regions (Fig. S5). We used data from ENCODE[38] to assess whether ASM is enriched in regions associated with DNase I hypersensitive (DHS) sites identified in multiple tissues including frontal cortex and cerebellum, as well as CD14+ monocytes and naïve B cells (as a proxy for blood). We compared the ASM score ranks for informative probes between regions defined by the presence or absence of DHS sites using a Wilcoxon rank-sum test (see **Materials and Methods**). DHS peaks across all tissues are enriched for higher ASM scores identified in cortex (BA9), cerebellum, and whole blood (Table S3 and Fig. S6). Of note, the most striking enrichment is found for cerebellum ASM scores in regions characterized by cerebellum DHS peaks in ENCODE ($P = 3.51 \times 10^{-220}$).

Inter-individual variation in ASM

We next examined inter-individual differences in ASM score at specific loci, defining probes with a large range of ASM scores across the 3 individuals as being characterized by "variable ASM." Differentially methylated regions (DMRs) associated with genomic imprinting, for example, are characterized by parental-origin-specific ASM and are expected to show consistently large ASM scores that exhibit allelic-flipping, resulting from genotype-independent ASM. Genotype-driven ASM, in contrast, is likely to be exemplified by consistent allelic biases in DNA methylation across individuals, and is generally not variable between individuals of the same genotype. Fig. S7 shows the correlation in ASM scores across the 3 individuals profiled by MSNP, with tissue-specific correlations given in Table S4. As expected, the individuals were more highly correlated for loci characterized by high ASM scores. For probes informative in at least 2 individuals we examined the range of

Table 3. Top 15 ASM sites in cortex (BA9), ASM score averaged across individuals.

Rank	SNP ID	Location	Associated gene(s)	BA9 ASM score	Cerebellum ASM score	Blood ASM score
1	SNP_A-8625237 (rs12493005)	3q26.32	*TBL1XR1*	0.24	0.24	0.20
2	SNP_A-8463467 (rs17164474)	7q35	*OR2F2, OR2F1*	0.23	0.17	0.17
3	SNP_A-1841543 (rs10234308)	7p15.3	*MGC87042*	0.23	0.25	0.23
4	SNP_A-8652129 (rs2479084)	1p36.21	*FHAD1*	0.22	0.13	0.18
5	SNP_A-8301602 (rs398225)	3p25.3	*SRGAP3, RAD18*	0.21	0.18	0.21
6	SNP_A-2107106 (rs987377)	6q21	*AIM1*[a], *ATG5*	0.20	0.16	0.15
7	SNP_A-2052542 (rs3098382)	5q13.2	*MAP1B*	0.19	0.24	0.04
8	SNP_A-2307481 (rs716591)	15q26.2	*LOC400456, MCTP2*	0.19	0.21	0.17
9	SNP_A-8643280 (rs3121125)	1q21.1	*HFE2, NBPF10*	0.19	0.17	0.15
10	SNP_A-8326632 (rs1542180)	2q31.1	*HOXD3, HOXD4*	0.19	0.24	0.21
11	SNP_A-1998023 (rs9722212)	9q34.11	*TOR1B, PTGES*	0.19	0.15	0.19
12	SNP_A-8640607 (rs12632177)	3q27.1	*MCF2L2*	0.18	0.11	0.07
13	SNP_A-4291638 (rs12670584)	7p13	*YKT6*	0.18	0.19	0.01
14	SNP_A-1892234 (rs17303015)	5p12	*MGC42105*	0.18	0.11	0.03
15	SNP_A-8653671 (rs4871852)	8p21.3	*TNFRSF10D*	0.18	0.09	0.11

[a] Known imprinted gene

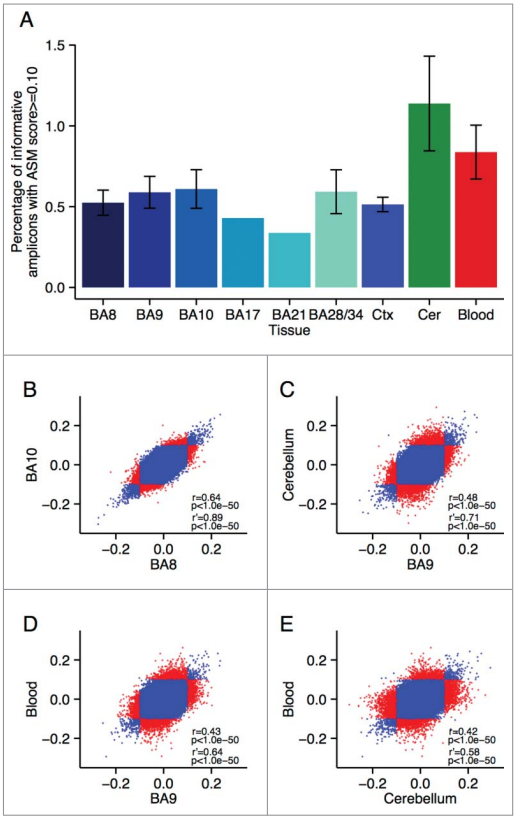


Figure 1. Allelic-skewing is less prevalent and less variable across cortical regions compared to cerebellum and whole blood. (A) The average proportion of informative amplicons showing an ASM score $\geq$ 0.10 for the 6 cortical regions profiled as well as averaged across the cortical areas (Ctx) is consistently lower (ASM score range = 0.34–0.61%) than in cerebellum (Cer) (ASM prevalence = 1.14%) or blood (ASM prevalence = 0.84%). Standard errors are shown for tissues for which samples were available from all 3 individuals. (B) – (E) Correlations of ASM scores are shown with each point representing one probe in one individual. Probes classified as allelically-skewed at an ASM score $\geq$ 0.10 in only one of the 2 compared tissues are highlighted in red. A higher degree of between-tissue variability is observed between cerebellum, cortex, and whole blood than between different cortical regions (shown as an example is BA8 vs. BA10). This difference becomes even more pronounced when restricting the set of probes to those that show allelic-skewing at an ASM score $\geq$ 0.10 in at least one of the 2 compared tissues (see subset correlation *r'*).

ASM scores across individuals and identified the top ranked variable ASM probes in each tissue (Table S5-S7) as well as cross-tissue variable sites, which show consistent inter-individual variation across all tissues (Table 6). Some sites show evidence of allelic-flipping in ASM score between individuals, indicative of genomic imprinting. These included several probes in the vicinity of the imprinted gene cluster on chromosome 15q11.2 (Figs. 4A-C). High ASM scores were also observed in the vicinity of other known imprinted loci, for example, *SNRPN* (Fig. S8A), *DLGAP2* (Fig. S8B), *AIM1* (Fig. S8C), *MEG3* (Fig. S8D), *BLCAP* (Fig. S8E), and *GRB10* (Fig. S8F), in addition to loci suspected to be imprinted, e.g., *TRAPPC9* (Fig. S8G), *EVX1* (Fig. S8H), and *TGFBI/VTRNA2* (Fig. S8I); however, we were unable to examine variable ASM in many of these regions because they were only informative (i.e., heterozygous) in a single individual. Notably, we also identified allelic-flipping in the vicinity of loci not previously characterized as being imprinted, for example *WRB* (Fig. 4D) and *ITPKI* (Fig. S9). Other variable ASM sites are marked by both high and low ASM scores in different individuals, rather than allelic-flipping between them, for example, *MGST3/LOC400794* (Table S5). Interestingly, we identified a number of sites characterized by tissue-specific variable ASM. A notable example is the imprinted gene *GRB10*, which has been previously shown to be differentially maternally- and paternally-expressed in a tissue-specific manner[29,33] (Fig. S8F).

Variable ASM sites are flanked by extended regions of intermediate DNA methylation

We next quantified genome-wide patterns of DNA methylation in a larger sample (n=39) of matched whole blood, cortex (BA9), and cerebellum samples using the Illumina Infinium HumanMethylation450 BeadChip (450K array). For the 100 top-ranked ASM sites in each of the 3 tissues, plus the 100 top-ranked cross-tissue, tissue-specific, and variable ASM sites we identified probes on the array located within 1 kb of the ASM marker SNPs (Table S8; detailed in Table S9–S17) to investigate patterns of DNA methylation across an extended region. As expected, regions around known imprinted loci identified by our ASM analysis are flanked by extended regions of intermediate DNA methylation (i.e., average levels of DNA methylation between 0.4 and 0.6) (Fig. 5A and Fig. 5B). We observe a highly significant enrichment (P range = 6.82×10^{-11} – 0.005) of intermediate DNA methylation relative to overall levels identified on the 450K array in regions flanking variable ASM sites in all 3 tissues (Table 7 and Fig. 6). For example, intermediate DNA methylation was observed around the variable ASM site overlapping *WRB* (Fig. 5C), that showed evidence of allelic-flipping (Fig. 4D). Another probe exhibiting variable ASM annotated to *TGFBI/VTRNA2-1* on chromosome 5 also shows a similar pattern of intermediate DNA methylation (Fig. 5D). Interestingly, 4 individuals are distinguished by consistent hypomethylation in whole blood and cortex across 16 of the 19 sites, consistent with previous reports describing polymorphic imprinting of this locus.[39,40]

Identification of tissue-specific genotype-driven ASM

In contrast to the intermediate DNA methylation patterns enriched in the vicinity of variable ASM sites, non-variable ASM (i.e., characterized by consistent ASM scores across individuals) is not significantly enriched for intermediate DNA methylation and in some cases is exemplified by trimodal patterns of DNA methylation, which are highly suggestive of genotype-driven ASM acting in *cis*. Of note, we also observe examples of tissue-specific genotype-driven ASM. For example, a tissue-specific ASM site identified as showing allelically-skewed DNA methylation in cerebellum (ASM score = 0.16) but not whole blood (ASM score = 0.03) or cortex (ASM score = 0.04) (Fig. 7A) located in an intron of the gene *SYNJ2* is

Table 4. Top 15 loci characterized by consistent ASM across cerebellum, whole blood, and cortex (BA9).

Rank	SNP ID	Location	Associated gene(s)	Cerebellum ASM score	Blood ASM score	BA9 ASM score	Tissue average
1	SNP_A-1841543 (rs10234308)	7p15.3	*MGC87042*	0.25	0.23	0.23	0.24
2	SNP_A-8625237 (rs12493005)	3q26.32	*TBL1XR1*	0.24	0.20	0.24	0.22
3	SNP_A-8326632 (rs1542180)	2q31.1	*HOXD3, HOXD4*	0.24	0.21	0.19	0.21
4	SNP_A-8301602 (rs398225)	3p25.3	*SRGAP3, RAD18*	0.18	0.21	0.21	0.20
5	SNP_A-8463467 (rs17164474)	7q35	*OR2F2, OR2F1*	0.17	0.17	0.23	0.19
6	SNP_A-2307481 (rs716591)	15q26.2	*MCTP2, LOC400456*	0.21	0.17	0.19	0.19
7	SNP_A-1894705 (rs986324)	Xp22.11	*DDX53, ZNF645*	0.18	0.20	0.17	0.18
8	SNP_A-2071005 (rs4687210)	3q28	*UTS2D*	0.16	0.22	0.17	0.18
9	SNP_A-1998023 (rs9722212)	9q34.11	*PTGES, TOR1B*	0.15	0.19	0.19	0.18
10	SNP_A-8652129 (rs2479084)	1p36.21	*FHAD1*	0.13	0.18	0.22	0.18
11	SNP_A-8696273 (rs1003533)	5q31.1	*C5orf56*	0.29	0.16	0.07	0.17
12	SNP_A-1862496 (rs17578280)	1p31.1	*LRRIQ3, NEGR1*	0.18	0.17	0.17	0.17
13	SNP_A-2040586 (rs17097827)	14q32.2	*C14orf177, BCL11B*	0.23	0.10	0.18	0.17
14	SNP_A-1932077 (rs220030)	15q11.2	*SNRPN*[a]	0.18	0.18	0.15	0.17
15	SNP_A-2107106 (rs987377)	6q21	*AIM1*[a], *ATG5*	0.16	0.15	0.20	0.17

[a] Known imprinted gene

flanked by trimodal levels of DNA methylation in cerebellum but not blood or cortex (Fig. 7B).

Discussion

This study confirms the relatively widespread distribution of allelically-skewed DNA methylation in the human genome, corroborating our previous data generated in whole blood.[7] We also present evidence for tissue-specific differences in the quantity and distribution of ASM between different regions of the human brain, and between brain and whole blood. Our findings are in line with previous reports, confirming the importance of tissue-specific DNA methylation profiles across the brain.[22,41]

Although our data confirm previous studies, identifying more between-tissue variation than inter-individual variation,[28] we find clear examples where ASM is variable between individuals. While the number of samples profiled in this study is too small to accurately determine how much of the observed inter-individual variation in ASM results from genetic and non-genetic effects, previous studies suggest that the majority of such variation is likely to be genetically driven.[7,8,11] Interestingly, we identify instances of tissue-specific allelically-skewed DNA methylation resulting from both genomic imprinting and genotypic effects. For example, we observe tissue-specific variable ASM around the imprinted growth factor receptor-bound protein 10 gene (*GRB10*), which encodes a protein that interacts with insulin-like growth factors[42,43] and we observe genotype-driven ASM exclusively in cerebellum for several probes within the synaptojanin 2 gene (*SYNJ2*), which encodes a protein involved in the uncoating of vesicles.[44,45] Such tissue-specific ASM has important implications for epigenetic epidemiology, and provides a mechanism by which genotype may exert an effect on gene function and regulation in a tissue-specific manner.

This study has a number of important limitations. First, although we used a unique set of samples comprising of matched tissues obtained from the same donors, the number of individuals profiled in our analysis was small, meaning we cannot definitively distinguish between genetic and non-genetic effects, or make broad statements about general patterns of inter-individual variation of ASM. Our Illumina 450K array validation studies were undertaken in a larger set of individuals, but could only confirm intermediate levels of DNA methylation and not detect allele-specific patterns. Second, given the limited availability of RNA from the same samples, we were unable to relate our ASM findings to allelic patterns of gene expression in the same individuals. Previous studies, however, have shown that ASM is linked to allele-specific expression of nearby genes.[28] Third, our analyses were undertaken on whole tissue, and represent aggregate values across a number of individual cell types. Fourth, the MSNP approach utilizes SNP

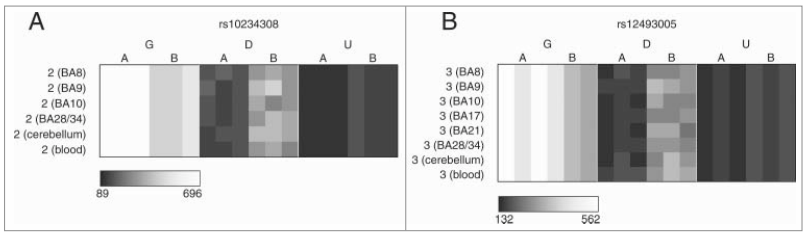


Figure 2. Multiple loci are characterized by consistent allelic-skewing of DNA methylation across all tissues. Heatmaps display allele signal intensities for genomic DNA (G), MSRE-digested DNA (D) and fully unmethylated, MSRE-digested DNA (U) in all tissues. A and B denote the 2 alleles of the SNP and brightness represents the quantile normalized signal intensity, with the scale displayed below the heatmap. Shown are the 2 top-ranked probes characterized by consistent ASM across tissues. These two probes were informative in (A) individual 2 for rs10234308 and (B) individual 3 for rs12493005.

Table 5. Top 15 tissue-specific ASM sites, defined by highly variable ASM scores across cerebellum, whole blood, and cortex (BA9).

Rank	SNP ID	Location	Associated gene (s)	Cerebellum ASM score	Blood ASM score	BA9 ASM score	SD
1	SNP_A-4255628 (rs959246)	18q12.3	*SLC14A2, SETBP1*	−0.30	−0.03	−0.05	0.15
2	SNP_A-1946136 (rs927000)	20q13.12	*STK4*	−0.04	0.23	0.01	0.15
3	SNP_A-8438077 (rs585451)	15q21.2	*ATP8B4, DTWD1*	0.01	−0.25	0.00	0.15
4	SNP_A-8397727 (rs1123514)	5q13.3	*ENC1, RGNEF*	−0.16	0.11	−0.01	0.13
5	SNP_A-2273834 (rs2252267)	14q23.1	*PRKCH*	0.19	−0.03	−0.05	0.13
6	SNP_A-4264458 (rs1358229)	4q27	*TRPC3, KIAA1109*	−0.23	0.01	0.00	0.13
7	SNP_A-8643616 (rs11700515)	21q22.3	*COL6A1, PCBP3*	0.20	−0.01	−0.03	0.13
8	SNP_A-2180729 (rs10276966)	7p15.2	*HIBADH, EVX1*[b]	−0.01	−0.25	−0.03	0.13
9	SNP_A-4210659 (rs10517764)	4q32.2	*NAF1, FSTL5*	0.07	−0.14	−0.15	0.13
10	SNP_A-1997061 (rs10512149)	9q21.33	*SLC28A3, NTRK2*	−0.19	0.02	0.05	0.13
11	SNP_A-2160121 (rs7205794)	16q23.3	*CDH13, MPHOSPH6*	−0.24	−0.01	−0.01	0.13
12	SNP_A-8667432 (rs519782)	1p36.11	*C1orf201*	0.19	−0.04	−0.01	0.13
13	SNP_A-1787058 (rs10951911)	7p12.3	*TNS3, C7orf65*	−0.20	−0.01	0.04	0.13
14	SNP_A-1931666 (rs6707698)	2q34	*IKZF2, ERBB4*	−0.22	−0.03	0.02	0.13
15	SNP_A-8502638 (rs10989120)	9q31.1	*TMEFF1, C9orf30*	0.04	−0.20	−0.15	0.13

[b] Suspected imprinted gene

microarrays—these do not interrogate the whole genome, and can only assess pools of DNA molecules. Allelic patterns of DNA methylation across individual DNA molecules cannot be directly assessed using this approach, as would be possible using bisulfite-sequencing methods. Furthermore, our threshold for calling allelic imbalances in DNA methylation is

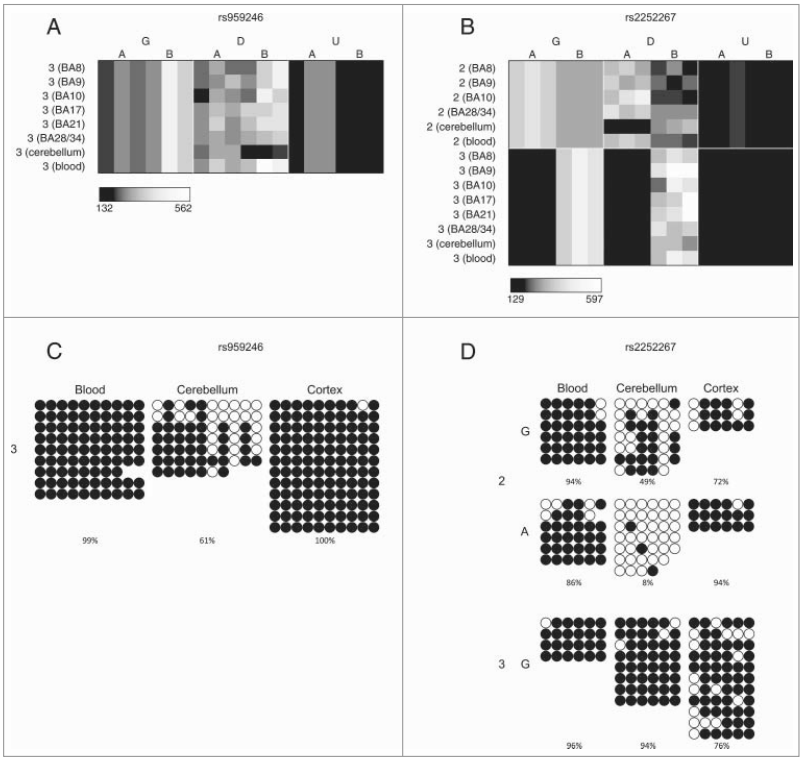


Figure 3. A number of loci are characterized by allelic-skewing of DNA methylation in only one tissue with minimal ASM present in any of the other tissues examined. Heatmaps display allele signal intensities for genomic DNA (G), MSRE-digested DNA (D) and fully unmethylated, MSRE-digested DNA (U) in all tissues. A and B denote the 2 alleles of the SNP and brightness represents the quantile normalized signal intensity, with the scale displayed below the heatmap. Two of the top-ranked cerebellum-specific ASM signals are (A) rs959246 (informative for individual 3) and (B) rs2252267 (informative for individual 2). The tissue-specific patterns of DNA methylation in these 2 loci were validated by clonal bisulfite sequencing, confirming the findings from the MSNP assay (C,D). Each row represents a single DNA molecule, with black dots depicting methylated cytosines and white dots depicting unmethylated cytosines. The percentage of methylated cytosines for each sample is displayed below the plots. The amplicon spanning the DMR associated with rs959246 in (C) did not encompass a SNP enabling us to distinguish between the 2 alleles, however the methylation pattern shows evidence for tissue-specific intermediate methylation (IM) in the cerebellum. G and A in (D) denote the 2 alleles determined by SNP variation within the amplicon in heterozygous individual 2. Cerebellum-specific hypomethylation of the A allele is observed surrounding rs2252267, with individual 3 (homozygous for the G allele) being highly methylated in all 3 tissues.

Table 6. Top 15 variable ASM sites, defined by average range of ASM scores between individuals across cerebellum, whole blood, and cortex (BA9).

Rank	SNP ID	Location	Associated gene(s)	Cerebellum range	Blood range	BA9 range	Average range
1	SNP_A-8579417 (rs1538116)	1q24.1	*MGST3, LOC400794*	0.24	0.35	0.32	0.30
2	SNP_A-2019421 (rs2244352)	21q22.2	*WRB*	0.36	0.24	0.29	0.30
3	SNP_A-4219174 (rs2346019)	5q31.1	*TGFBI, VTRNA2*[b]	0.23	0.23	0.30	0.25
4	SNP_A-4208914 (rs927651)	20q13.2	*CYP24A1*	0.25	0.23	0.17	0.22
5	SNP_A-8717059 (rs12713666)	2p13.3	*ARHGAP25*	0.26	0.15	0.21	0.21
6	SNP_A-8692937 (rs4605656)	4q24	*CXXC4, TACR3*	0.16	0.18	0.25	0.20
7	SNP_A-8634251 (rs1209228)	14q24.2	*RGS6, SIPA1L1*	0.19	0.15	0.24	0.19
8	SNP_A-8329713 (rs16825906)	3q13.31	*LSAMP, IGSF11, LOC285194*	0.21	0.13	0.22	0.19
9	SNP_A-8713358 (rs16890883)	4p15.33	*CPEB2, LOC152742*	0.24	0.14	0.17	0.19
10	SNP_A-8638348 (rs4525744)	2p21	*SRBD1*	0.20	0.18	0.17	0.18
11	SNP_A-1855770 (rs3764124)	13q34	*CUL4A*	0.16	0.23	0.15	0.18
12	SNP_A-4259064 (rs7766133)	6p22.3	*MBOAT1*	0.17	0.17	0.20	0.18
13	SNP_A-2118217 (rs1695824)	1p36.33	*VWA1, TMEM88B*	0.25	0.16	0.13	0.18
14	SNP_A-1880775 (rs6116750)	20p12.3	*PROKR2*	0.23	0.13	0.18	0.18
15	SNP_A-8424056 (rs3922835)	18q12.1	*CDH2*[b], *CHST9*	0.18	0.04	0.32	0.18

[b] Suspected imprinted gene

somewhat arbitrary; it is likely that our data is confounded by both false positives and negatives. We did, however, find very consistent overlap in whole blood ASM data with that reported in our previous study using the same laboratory and analysis methods,[7] confirming the validity of the MSNP approach. Furthermore, we validated our findings using 2 independent platforms: clonal bisulfite sequencing and the Illumina 450K Human methylation array. Using the latter, we were able to show that variable ASM sites are located in an extended context of intermediate DNA methylation, supporting a regional regulatory role of DNA methylation in these domains, which is potentially driving intermediate expression levels in a quantitative manner across gene regulation clusters.[12] In addition, we observed a significant enrichment of ASM in regions

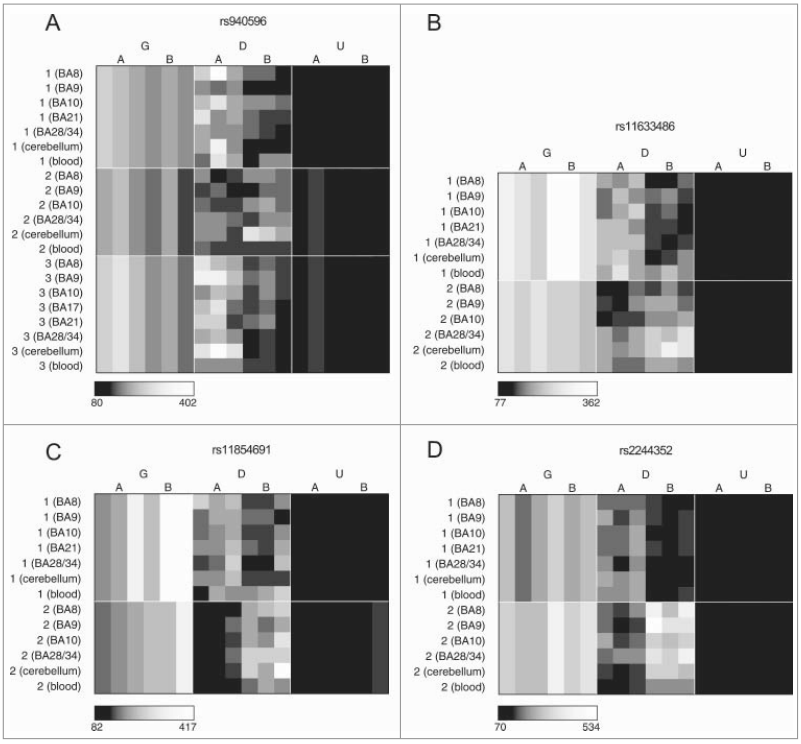


Figure 4. Cases of allelic-flipping of DNA methylation between individuals are found both in known imprinted gene clusters as well as regions not previously confirmed as being imprinted. The heatmaps show allele signal intensities for genomic DNA (G), MSRE-digested DNA (D) and fully unmethylated, MSRE-digested DNA (U) in all tissues. A and B denote the 2 alleles of the SNP and brightness represents the quantile normalized signal intensity, with the scale displayed below the heatmap. (A)-(C) Three probes in the vicinity of the known imprinted cluster on chromosome 15q11.2 show variable ASM and allelic-flipping in cerebellum [(A) rs940596, (B) rs11633486, (C) rs11854691]. (D) Allelic-flipping of DNA methylation across all tissues is observed in the vicinity of WRB (rs2244352), which has not been reported as imprinted previously.

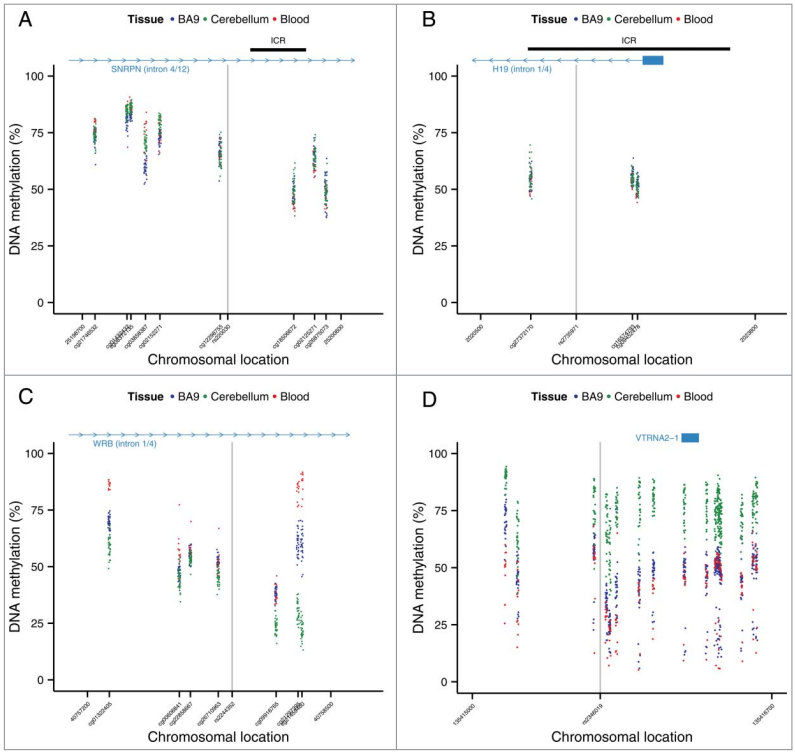


Figure 5. Regions around variable ASM sites are enriched for genomic domains characterized by intermediate DNA methylation. Flanking regions (1 kb) of variable ASM sites show a significant enrichment in intermediately methylated probes on the Illumina 450K Human Methylation Array (P range = 6.82×10^{-11} – 0.005). The scatter plots (A)-(D) show the location of genes and imprinting control regions (ICR) if overlapping the plotting window, as well as the location of the SNP from the MSNP assay (gray vertical line). These intermediate DNA methylation patterns span several known imprinted regions, for example (A) a flanking region of rs220030 in the *SNRPN* imprinted DMR, and (B) a region around rs2735971 in the imprinted gene *H19*. (C) Other sites showing allelic-flipping and intermediately methylated flanking regions lie in areas not previously known to be imprinted, for example rs2244352, which lies in an intron of *WRB*. (D) A polymorphic ASM pattern is observed in a flanking region around rs2346019, a downstream gene variant *of VTRNA2-1*, in which the majority of samples display intermediate DNA methylation in cortex (BA9) and whole blood. Of note, 4 samples show consistent hypomethylation across this region.

characterized by DHS peaks across several tissues. This enrichment of ASM in the vicinity of markers of open chromatin supports the involvement of ASM in transcriptional activity.

To conclude, we explored inter- and intra-individual variation in ASM across several regions of the human brain and whole blood from multiple individuals. Consistent with previous studies, we find relatively widespread ASM, observing allelically-skewed DNA methylation flanking known imprinted regions, and show that ASM sites are often located in an extended genomic context of intermediate DNA methylation. Interestingly, we detect cases of genotype-driven ASM, which are also tissue-specific. These findings contribute to our understanding about the nature of differential DNA methylation across tissues and have important implications for genetic studies of complex disease. As a resource to the community, ASM patterns across each of the tissues studied are available in a searchable online database: http://epigenetics.essex.ac.uk/ASMBrainBlood.

Table 7. Enrichment in intermediate methylation (IM) near ASM regions.

	IM rate	O/E	*P*
Blood[d]	5.88%	0.98	0.576
Cerebellum[d]	11.49%	1.75	0.059
BA9[d]	13.29%	1.45	0.065
Cross-tissue	8.11%	0.87	0.741
Tissue-specific	8.00%	0.85	0.701
Variable	38.46%	4.10	9.56×10^{-11}
Variable (BA9)[d]	37.50%	4.09	8.86×10^{-9}
Variable (cerebellum)[d]	16.95%	2.58	0.005
Variable (blood)[d]	34.48%	5.77	6.82×10^{-11}

[c] IM is defined as an average methylation between 0.4 and 0.6, enrichment *P* values are based on a hypergeometric test based on the background distribution of IM.
[d] We used tissue-specific background distributions for ASM types based on single tissues.

Materials and methods

Genome-wide analysis of allelically-skewed DNA methylation

Post-mortem brain and pre-mortem whole blood samples from 2 female and one male donors were provided by the MRC London Neurodegenerative Disease Brain Bank (http://www.kcl.ac.uk/ioppn/depts/bcn/Our-research/Neurodegeneration/brain-bank.aspx). Subjects were approached in life for written consent for

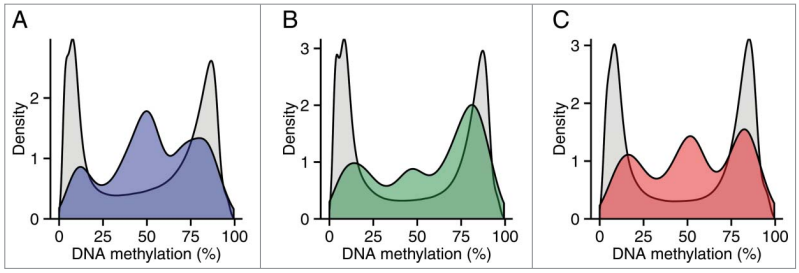


Figure 6. Regions around variable ASM are enriched in intermediate DNA methylation. The distributions of DNA methylation at the 65 Illumina 450K Human Methylation Array probes within 1 kb of the top 100 variable ASM sites (Table S14) show an enrichment in intermediately methylated probes compared to DNA methylation levels across the whole array (shown in gray) in (A) cortex (BA9), (B) cerebellum, and (C) whole blood).

brain banking, and all tissue donations were collected and stored following legal and ethical guidelines (NHS reference number 08/MRE09/38; the HTA license number for the LBBND brain bank is 12293). All samples were free from neuropathological and neuropsychiatric disease. A detailed list of brain regions obtained for each individual is provided in Table S18. Genomic DNA was isolated from all tissue samples using a standard phenol-chloroform protocol and assessed for purity and degradation prior to analysis (see Davies et al.[22] for additional information about the samples used in this study). The MSNP method, described previously,[6,7] was used to quantitatively assess allelic-skewing of DNA methylation across the genome. Briefly, Affymetrix Genome-wide Human SNP 6.0 Arrays were used to genotype a) DNA from each tissue sample digested with a cocktail of MSREs (HpaII: 5′ -C C G G-3′, HhaI: 5′ -G C G C-3′, and AciI: 5′ -C^C G C-3′) (D arrays), b) unmethylated whole-genome-amplified DNA for each individual digested with the same cocktail of MSREs to control for possible confounding effects of DNA sequence polymorphisms located at MSRE cut-sites (U arrays), and c) genomic DNA from each of the 3 individuals to identify heterozygous (informative) SNPs (G arrays). Unmethylated DNA was produced by whole-genome amplifying 100 ng cerebellum DNA using the Qiagen RepliG kit (Qiagen, Crawley, UK) using the manufacturer's protocol. In total 28 genotyping arrays were processed: 22 D arrays (DNA from between 6 and 7 brain regions plus whole blood, for each individual), 3 U arrays (one for each individual), and 3 G arrays (one for each individual). Additional methodological details are available in Schalkwyk et al.[7]

Selection of informative SNPs and quantification of ASM

To be informative in the ASM assay, SNPs must be heterozygous, and the amplicon must contain an MSRE cut site.[6,7] To guard against poorly performing SNP probes we also removed consistently low signal intensity SNPs across the 22 G arrays, and those yielding a highly variable U/G signal ratio (SD > 0.077) across all samples. A total of 220,449 SNPs passed our stringent filtering criteria and were classified as informative and heterozygous in at least one individual. The number of informative SNPs in each of the individual samples profiled by the MSNP method is shown in Table S19. Quantitative measures of ASM were derived by comparing signal intensities between the D (MSRE digested) and G (genomic DNA) arrays using the SNPMaP package (v1.02) in R that was developed for the estimation of allele frequencies in DNA pools genotyped on SNP arrays.[46] Briefly, relative allele score (RAS) values were generated for all SNPs on the array, which are defined as A/(A + B), where A and B are the intensities of the probes for the 2 alleles of

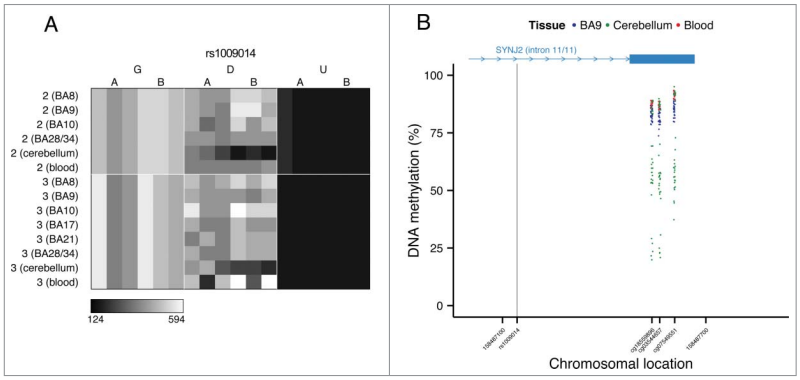


Figure 7. Genotype-driven ASM can be tissue-specific. (A) The tissue-specific ASM site rs1009014, located in an intron of *SYNJ2*, shows allelic-skewing of DNA methylation in cerebellum (ASM score = 0.16) but not in cortex (ASM score = 0.04) or whole blood (ASM score = 0.03). (B) DNA methylation levels for Illumina 450K Human Methylation Array probes in cerebellum across a flanking region of this locus exhibit a genotype-driven tissue-specific DNA methylation pattern. The scatter plot shows the location of *SYNJ2* (transcript variant 1), as well as the location of the informative SNP from the MSNP assay (gray vertical line).

a given SNP. For a given SNP in a heterozygous individual, ASM (or allelic-skewing of DNA methylation) is detected as a difference in RAS between the G and D arrays. We call this difference in RAS "ASM score" and define probes showing an absolute ASM score $\geq$ 0.10 as "allelically-skewed." A UCSC custom annotation track showing the location of all 220,449 loci and the degree of allelic-skewing in DNA methylation across each tissue and individual is available for download from our website (http://epigenetics.essex.ac.uk/ASMBrainBlood). Enrichment analyses were performed using a Kruskal-Wallis rank-sum test for ASM rank differences between the annotated genic regions. This non-parametric method tests whether multiple samples were drawn from the same distribution and is the multivariate extension of the better-known Wilcoxon rank-sum test. This test allowed us to avoid selecting a specific threshold for ASM scores and does not assume a normal distribution of residuals. Of the 220,449 informative probes, 219,921 could be annotated to specific defined genic regions. Annotations were based on the *Homo sapiens* hg19 build from UCSC using the AnnotationHub Bioconductor package[47] classifying probes as residing in introns (n = 100,254), 5′UTRs (n = 347), 3′UTRs (n = 2,341), coding regions (n = 2,569), intergenic regions (n = 110,186) and promoters (n = 4,224). A Nemenyi test for pairwise multiple comparisons of mean rank sums as implemented in the PMCMR R package[48] was used for post-hoc comparisons. ENCODE tracks for DHS peaks in frontal cortex, frontal cerebrum, cerebellum, CD14+ monocytes, naïve B cells, H1 human embryonic stem cells (H1-hesc), heart, and fibroblasts were obtained from the UCSC genome browser (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeOpenChromDnase/). For the DHS enrichment analyses we tested whether the rank-sums for ASM scores differed significantly in informative probes defined by the presence or absence of DHS peaks using a Wilcoxon rank-sum test. Informative probes were ranked according to ASM scores with higher absolute ASM scores corresponding to lower ranks.

Clonal bisulfite sequencing

Two regions were subsequently selected for clonal bisulfite sequencing analysis to further verify our findings and determine the precise allele-specific patterns of DNA methylation. Following sodium bisulfite treatment and bisulfite-PCR amplification, amplicons were cloned using the TOPO TA cloning method (Invitrogen, Paisley, UK) and sequenced with BigDye v1.1 sequencing chemistry (Applied Biosystems) (Table S20). Sequencing traces were visualized, quality controlled, and aligned using BiQ Analyzer.[49] All data were tested for complete sodium bisulfite conversion, with an overall conversion rate > 99.9% estimated by BiQ Analyzer.

Validation of ASM on the Illumina 450K HumanMethylation microarray

Further analysis of ASM sites was undertaken on a larger collection of post-mortem brain samples (n = 34), comprising BA9, BA21, BA28/34, and cerebellum, which were also free of any neuropathology and neuropsychiatric disease. Additionally we analyzed matched pre-mortem whole blood samples, which were available for a subset (n = 8), as well as 5 unmatched blood samples (see Table S18). The 3 individuals profiled by MSNP were included in this analysis. DNA (500 ng) from each sample was treated with sodium bisulfite in duplicate, using the EZ-96 DNA methylation kit (Zymo Research, CA, USA). DNA methylation was quantified using the Illumina Infinium HumanMethylation450 BeadChip (Illumina Inc., CA, USA) run on an Illumina HiScan System (Illumina) using the manufacturers' standard protocol, with pre-processing and stringent quality control performed as previously described.[50] We used the GenomicRanges package[51] to extract data for all CpG sites within 1 kb of candidate ASM SNPs and examined patterns of DNA methylation across the 3 tissues. Intermediate DNA methylation was defined as an average methylation value between 0.4 and 0.6 across all individuals. To test for statistical significance of enrichment in intermediately methylated probes we used a hypergeometric distribution based on the number of probes tested and the background of intermediately methylated probes across the whole array. Annotation of genes in the methylation plots (Fig. 5 and Fig. 7B) was obtained from the UCSC Genome Browser hg19 assembly. Imprinting control region (ICR) annotation was obtained from the web resource on human DMRs provided by the Department of Medical and Molecular Genetics, Kings College London (https://atlas.genetics.kcl.ac.uk) and lifted over from hg18 to hg19.

Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

Acknowledgments

This work was supported by grants from the UK Medical Research Council (MRC) (grant number MR/K013807/1) and US National Institutes of Health (grant number AG036039) to JM. SJM is funded by the EU-FP7 project EpiTrain (REA grant agreement n° 316758).

References

1. Ferguson-Smith AC, Sasaki H, Cattanach BM, Surani MA. Parental-origin-specific epigenetic modification of the mouse H19 gene. Nature 1993; 362:751-5; PMID:8469285; http://dx.doi.org/10.1038/362751a0
2. Wutz A, Smrzka OW, Schweifer N, Schellander K, Wagner EF, Barlow DP. Imprinted expression of the Igf2r gene depends on an intronic CpG island. Nature 1997; 389:745-9; PMID:9338788; http://dx.doi.org/10.1038/39631
3. Lee JT, Bartolomei MS. X-inactivation, imprinting, and long noncoding RNAs in health and disease. Cell 2013; 152:1308-23; PMID:23498939; http://dx.doi.org/10.1016/j.cell.2013.02.016
4. Ferguson-Smith AC. Genomic imprinting: the emergence of an epigenetic paradigm. Nat Rev Genet 2011; 12:565-75; PMID:21765458; http://dx.doi.org/10.1038/nrg3032
5. Barlow DP, Bartolomei MS. Genomic imprinting in mammals. Cold Spring Harb Perspect Biol 2014; 6:pii: a018382; PMID:24492710; http://dx.doi.org/10.1101/cshperspect.a018382
6. Kerkel K, Spadola A, Yuan E, Kosek J, Jiang L, Hod E, Li K, Murty VV, Schupf N, Vilain E, et al. Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. Nat Genet 2008; 40:904-8; PMID:18568024; http://dx.doi.org/10.1038/ng.174
7. Schalkwyk LC, Meaburn EL, Smith R, Dempster EL, Jeffries AR, Davies MN, Plomin R, Mill J. Allelic skewing of DNA methylation is

widespread across the genome. Am J Hum Genet 2010; 86:196-212; PMID:20159110; http://dx.doi.org/10.1016/j.ajhg.2010.01.014
8. Shoemaker R, Deng J, Wang W, Zhang K. Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. Genome Res 2010; 20:883-9; PMID:20418490; http://dx.doi.org/10.1101/gr.104695.109
9. Tycko B. Allele-specific DNA methylation: beyond imprinting. Hum Mol Genet 2010; 19:R210-20; PMID:20855472; http://dx.doi.org/10.1093/hmg/ddq376
10. Gertz J, Varley KE, Reddy TE, Bowling KM, Pauli F, Parker SL, Kucera KS, Willard HF, Myers RM. Analysis of DNA methylation in a three-generation family reveals widespread genetic influence on epigenetic regulation. PLoS genetics 2011; 7:e1002228; PMID:21852959; http://dx.doi.org/10.1371/journal.pgen.1002228
11. Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, Degner JF, Gilad Y, Pritchard JK. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. Genome biology 2011; 12:R10; PMID:21251332; http://dx.doi.org/10.1186/gb-2011-12-1-r10
12. Elliott G, Hong C, Xing X, Zhou X, Li D, Coarfa C, Bell RJ, Maire CL, Ligon KL, Sigaroudinia M, et al. Intermediate DNA methylation is a conserved signature of genome regulation. Nat Commun 2015; 6:6363; PMID:25691127; http://dx.doi.org/10.1038/ncomms7363
13. Reik W. Stability and flexibility of epigenetic gene regulation in mammalian development. Nature 2007; 447:425-32; PMID:17522676; http://dx.doi.org/10.1038/nature05918
14. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 2008; 454:766-70; PMID:18600261; http://dx.doi.org/10.1038/nature07107
15. Seisenberger S, Andrews S, Krueger F, Arand J, Walter J, Santos F, Popp C, Thienpont B, Dean W, Reik W. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol Cell 2012; 48:849-62; PMID:23219530; http://dx.doi.org/10.1016/j.molcel.2012.11.001
16. Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, Lucero J, Huang Y, Dwork AJ, Schultz MD, et al. Global epigenomic reconfiguration during mammalian brain development. Science 2013; 341:1237905; PMID:23828890; http://dx.doi.org/10.1126/science.1237905
17. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 2009; 462:315-22; PMID:19829295; http://dx.doi.org/10.1038/nature08514
18. Ziller MJ, Gu H, Muller F, Donaghey J, Tsai LT, Kohlbacher O, De Jager PL, Rosen ED, Bennett DA, Bernstein BE, et al. Charting a dynamic DNA methylation landscape of the human genome. Nature 2013; 500:477-81; PMID:23925113; http://dx.doi.org/10.1038/nature12433
19. Varley KE, Gertz J, Bowling KM, Parker SL, Reddy TE, Pauli-Behn F, Cross MK, Williams BA, Stamatoyannopoulos JA, Crawford GE, et al. Dynamic DNA methylation across diverse human cell lines and tissues. Genome Res 2013; 23:555-67; PMID:23325432; http://dx.doi.org/10.1101/gr.147942.112
20. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. Integrative analysis of 111 reference human epigenomes. Nature 2015; 518:317-30; PMID:25693563; http://dx.doi.org/10.1038/nature14248
21. Ladd-Acosta C, Pevsner J, Sabunciyan S, Yolken RH, Webster MJ, Dinkins T, Callinan PA, Fan JB, Potash JB, Feinberg AP. DNA methylation signatures within the human brain. American journal of human genetics 2007; 81:1304-15; PMID:17999367; http://dx.doi.org/10.1086/524110
22. Davies MN, Volta M, Pidsley R, Lunnon K, Dixit A, Lovestone S, Coarfa C, Harris RA, Milosavljevic A, Troakes C, et al. Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. Genome Biology 2012; 13:R43; PMID:22703893; http://dx.doi.org/10.1186/gb-2012-13-6-r43
23. Takizawa T, Nakashima K, Namihira M, Ochiai W, Uemura A, Yanagisawa M, Fujita N, Nakao M, Taga T. DNA methylation is a critical cell-intrinsic determinant of astrocyte differentiation in the fetal brain. Developmental cell 2001; 1:749-58; PMID:11740937; http://dx.doi.org/10.1016/S1534-5807(01)00101-0
24. Kozlenkov A, Roussos P, Timashpolsky A, Barbu M, Rudchenko S, Bibikova M, Klotzle B, Byne W, Lyddon R, Di Narzo AF, et al. Differences in DNA methylation between human neuronal and glial cells are concentrated in enhancers and non-CpG sites. Nucleic Acids Res 2014; 42:109-27; PMID:24057217; http://dx.doi.org/10.1093/nar/gkt838
25. Ziller MJ, Edri R, Yaffe Y, Donaghey J, Pop R, Mallard W, Issner R, Gifford CA, Goren A, Xing J, et al. Dissecting neural differentiation regulatory networks through epigenetic footprinting. Nature 2015; 518:355-9; PMID:25533951; http://dx.doi.org/10.1038/nature13990
26. Mo A, Mukamel EA, Davis FP, Luo C, Henry GL, Picard S, Urich MA, Nery JR, Sejnowski TJ, Lister R, et al. Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. Neuron 2015; 86:1369-84; PMID:26087164; http://dx.doi.org/10.1016/j.neuron.2015.05.018
27. Paliwal A, Temkin AM, Kerkel K, Yale A, Yotova I, Drost N, Lax S, Nhan-Chang CL, Powell C, Borczuk A, et al. Comparative anatomy of chromosomal domains with imprinted and non-imprinted allele-specific DNA methylation. PLoS genetics 2013; 9:e1003622; PMID:24009515; http://dx.doi.org/10.1371/journal.pgen.1003622
28. Schultz MD, He Y, Whitaker JW, Hariharan M, Mukamel EA, Leung D, Rajagopal N, Nery JR, Urich MA, Chen H, et al. Human body epigenome maps reveal noncanonical DNA methylation variation. Nature 2015; 523(7559):212-6; PMID:26605523; http://dx.doi.org/10.1038/nature16179
29. Prickett AR, Oakey RJ. A survey of tissue-specific genomic imprinting in mammals. Mol Genet Genomics 2012; 287:621-30; PMID:22821278; http://dx.doi.org/10.1007/s00438-012-0708-6
30. Mancini-Dinardo D, Steele SJ, Levorse JM, Ingram RS, Tilghman SM. Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. Genes Dev 2006; 20:1268-82; PMID:16702402; http://dx.doi.org/10.1101/gad.1416906
31. Kelsey G. Imprinting on chromosome 20: tissue-specific imprinting and imprinting mutations in the GNAS locus. Am J Med Genet C Semin Med Genet 2010; 154C:377-86; PMID:20803660; http://dx.doi.org/10.1002/ajmg.c.30271
32. Williamson CM, Ball ST, Dawson C, Mehta S, Beechey CV, Fray M, Teboul L, Dear TN, Kelsey G, Peters J. Uncoupling antisense-mediated silencing and DNA methylation in the imprinted Gnas cluster. PLoS genetics 2011; 7:e1001347; PMID:21455290; http://dx.doi.org/10.1371/journal.pgen.1001347
33. Arnaud P, Monk D, Hitchins M, Gordon E, Dean W, Beechey CV, Peters J, Craigen W, Preece M, Stanier P, et al. Conserved methylation imprints in the human and mouse GRB10 genes with divergent allelic expression suggests differential reading of the same mark. Hum Mol Genet 2003; 12:1005-19; PMID:12700169; http://dx.doi.org/10.1093/hmg/ddg110
34. Grundberg E, Meduri E, Sandling JK, Hedman AK, Keildson S, Buil A, Busche S, Yuan W, Nisbet J, Sekowska M, et al. Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. Am J Hum Genet 2013; 93:876-90; PMID:24183450; http://dx.doi.org/10.1016/j.ajhg.2013.10.004
35. Mill J, Heijmans BT. From promises to practical strategies in epigenetic epidemiology. Nat Rev Genet 2013; 14:585-94; PMID:23817309; http://dx.doi.org/10.1038/nrg3405
36. Meaburn EL, Schalkwyk LC, Mill J. Allele-specific methylation in the human genome: implications for genetic studies of complex disease. Epigenetics 2010; 5:578-82; PMID:20716955; http://dx.doi.org/10.4161/epi.5.7.12960
37. Hutchinson JN, Raj T, Fagerness J, Stahl E, Viloria FT, Gimelbrant A, Seddon J, Daly M, Chess A, Plenge R. Allele-specific methylation occurs at genetic variants associated with complex disease. PloS one

2014; 9:e98464; PMID:24911414; http://dx.doi.org/10.1371/journal.pone.0098464
38. Encode Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 2012; 489:57-74; PMID:22955616; http://dx.doi.org/10.1038/nature11247
39. Treppendahl MB, Qiu X, Sogaard A, Yang X, Nandrup-Bus C, Hother C, Andersen MK, Kjeldsen L, Mollgard L, Hellstrom-Lindberg E, et al. Allelic methylation levels of the noncoding VTRNA2-1 located on chromosome 5q31.1 predict outcome in AML. Blood 2012; 119:206-16; PMID:22058117; http://dx.doi.org/10.1182/blood-2011-06-362541
40. Romanelli V, Nakabayashi K, Vizoso M, Moran S, Iglesias-Platas I, Sugahara N, Simon C, Hata K, Esteller M, Court F, et al. Variable maternal methylation overlapping the nc886/vtRNA2-1 locus is locked between hypermethylated repeats and is frequently altered in cancer. Epigenetics 2014; 9:783-90; PMID:24589629; http://dx.doi.org/10.4161/epi.28323
41. Illingworth RS, Gruenewald-Schneider U, De Sousa D, Webb S, Merusi C, Kerr AR, James KD, Smith C, Walker R, Andrews R, et al. Inter-individual variability contrasts with regional homogeneity in the human brain DNA methylome. Nucleic Acids Res 2015; 43:732-44; PMID:25572316; http://dx.doi.org/10.1093/nar/gku1305
42. Holt LJ, Siddle K. Grb10 and Grb14: enigmatic regulators of insulin action–and more? Biochem J 2005; 388:393-406; PMID:15901248; http://dx.doi.org/10.1042/BJ20050216
43. Garfield AS, Cowley M, Smith FM, Moorwood K, Stewart-Cox JE, Gilroy K, Baker S, Xia J, Dalley JW, Hurst LD, et al. Distinct physiological and behavioural functions for parental alleles of imprinted Grb10. Nature 2011; 469:534-8; PMID:21270893; http://dx.doi.org/10.1038/nature09651
44. Rusk N, Le PU, Mariggio S, Guay G, Lurisci C, Nabi IR, Corda D, Symons M. Synaptojanin 2 functions at an early step of clathrin-mediated endocytosis. Curr Biol 2003; 13:659-63; PMID:12699622; http://dx.doi.org/10.1016/S0960-9822(03)00241-0
45. Verstreken P, Koh TW, Schulze KL, Zhai RG, Hiesinger PR, Zhou Y, Mehta SQ, Cao Y, Roos J, Bellen HJ. Synaptojanin is recruited by endophilin to promote synaptic vesicle uncoating. Neuron 2003; 40:733-48; PMID:14622578; http://dx.doi.org/10.1016/S0896-6273(03)00644-5
46. Davis OS, Plomin R, Schalkwyk LC. The SNPMaP package for R: a framework for genome-wide association using DNA pooling on microarrays. Bioinformatics 2009; 25:281-3; PMID:19008252; http://dx.doi.org/10.1093/bioinformatics/btn587
47. Morgan M, Carlson M, Tenenbaum D, Arora S. *AnnotationHub: Client to access AnnotationHub resources*. R package version 2.2.2.
48. Pohlert T (2014). *The Pairwise Multiple Comparison of Mean Ranks Package (PMCMR)*. R package, http://CRAN.R-project.org/package=PMCMR.
49. Bock C, Reither S, Mikeska T, Paulsen M, Walter J, Lengauer T. BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. Bioinformatics 2005; 21:4067-8; PMID:16141249; http://dx.doi.org/10.1093/bioinformatics/bti652
50. Pidsley R, CC YW, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. BMC genomics 2013; 14:293; PMID:23631413; http://dx.doi.org/10.1186/1471-2164-14-293
51. Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. Software for computing and annotating genomic ranges. PLoS Comput Biol 2013; 9:e1003118; PMID:23950696; http://dx.doi.org/10.1371/journal.pcbi.1003118

4.2. Supplementary Material

Supplementary Figure 1

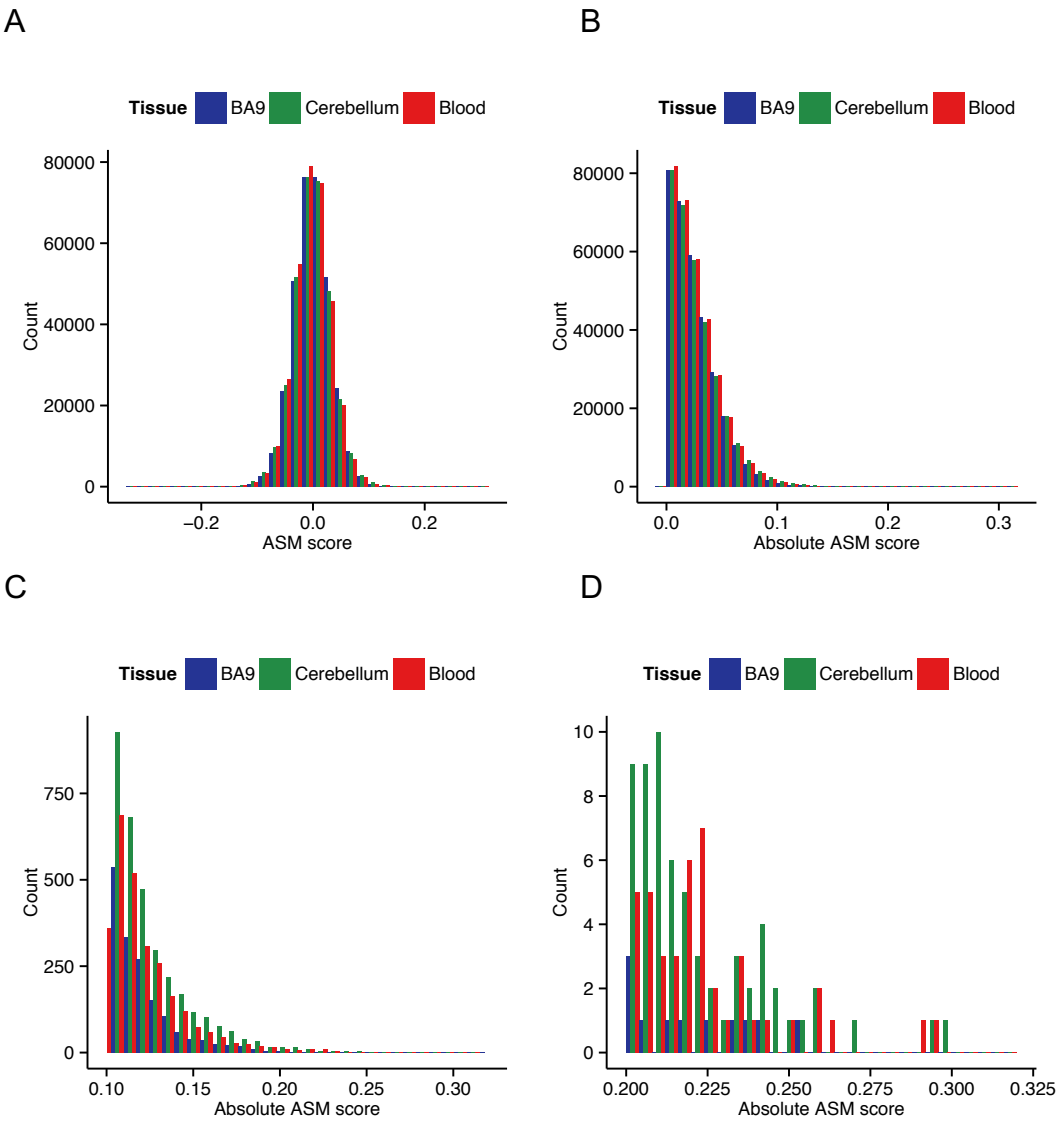


Figure 4-8. The overall distribution of ASM score is similar across brain regions and whole blood samples. The distribution of ASM scores (**A**) and absolute ASM scores (**B**) is similar across cortex (BA9), cerebellum and whole blood. However, histograms of absolute ASM score using a threshold of ≥ 0.10 (**C**) and ≥ 0.20 (**D**) indicate a large number of loci characterized by high allelic-skewing of DNA methylation in cerebellum and whole blood compared to cortex (BA9).

Supplementary Figure 2

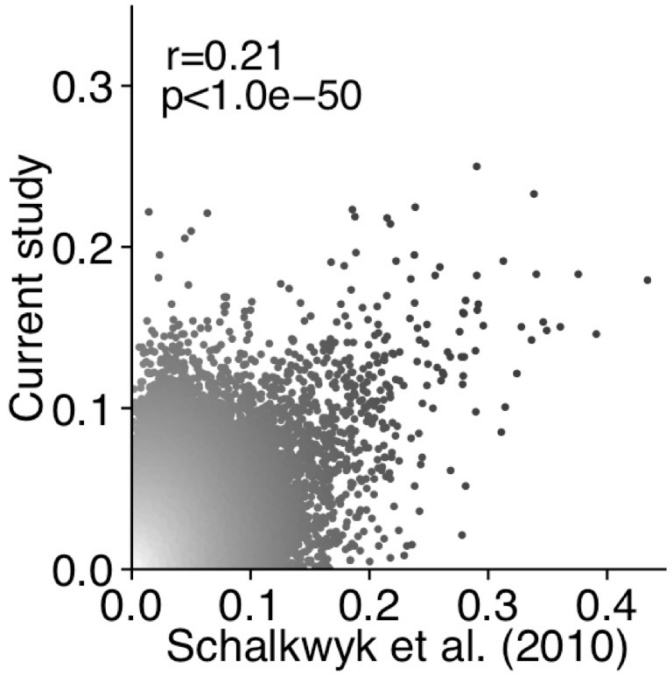


Figure 4-9. ASM scores in whole blood overlap with those identified in a previous study. Across probes informative in both studies there is a significant correlation between whole blood ASM scores identified in the current study and previously by Schalkwyk *et al* (2010) (r = 0.21, $P < 1.0 \times 10^{-50}$). 9 of the 15 top-ranked blood ASM sites in our current study (**Table 1**) were informative in our previous dataset; 7 of these were characterized by an absolute ASM score ≥ 0.10.

Supplementary Figure 3

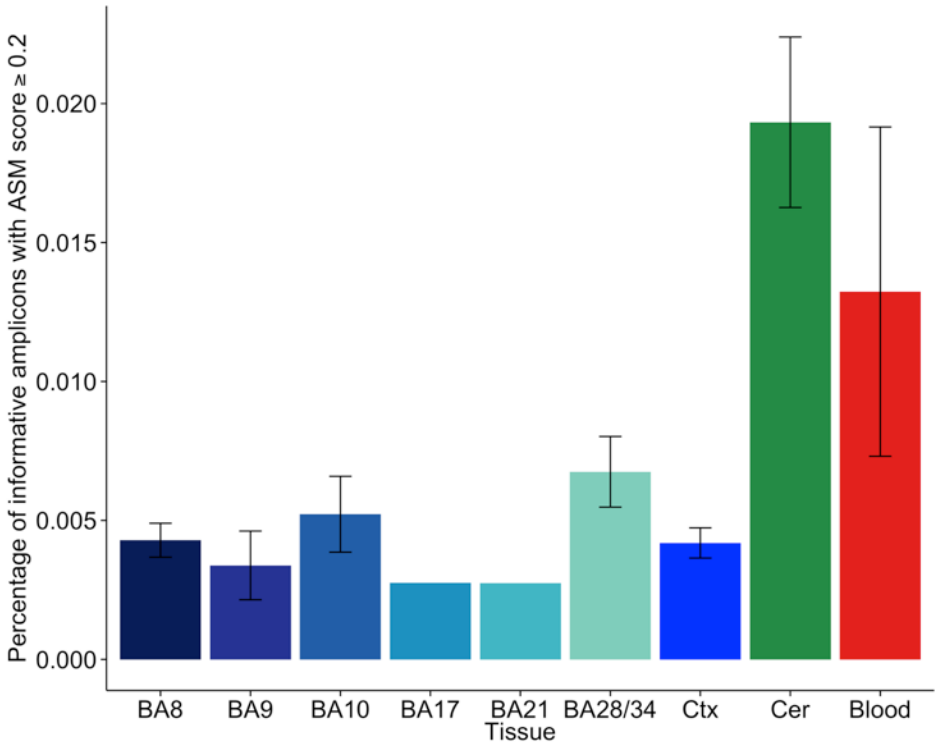


Figure 4-10. Large allelic asymmetries in DNA methylation are more prevalent in cerebellum and whole blood than cortex. The average percentage of informative amplicons showing an ASM score ≥ 0.20 for the six cortical regions profiled as well as averaged across the cortical areas (Ctx) is consistently lower than in cerebellum (Cer) or blood. Standard errors are shown for tissues for which samples were available from all three individuals.

Supplementary Figure 4

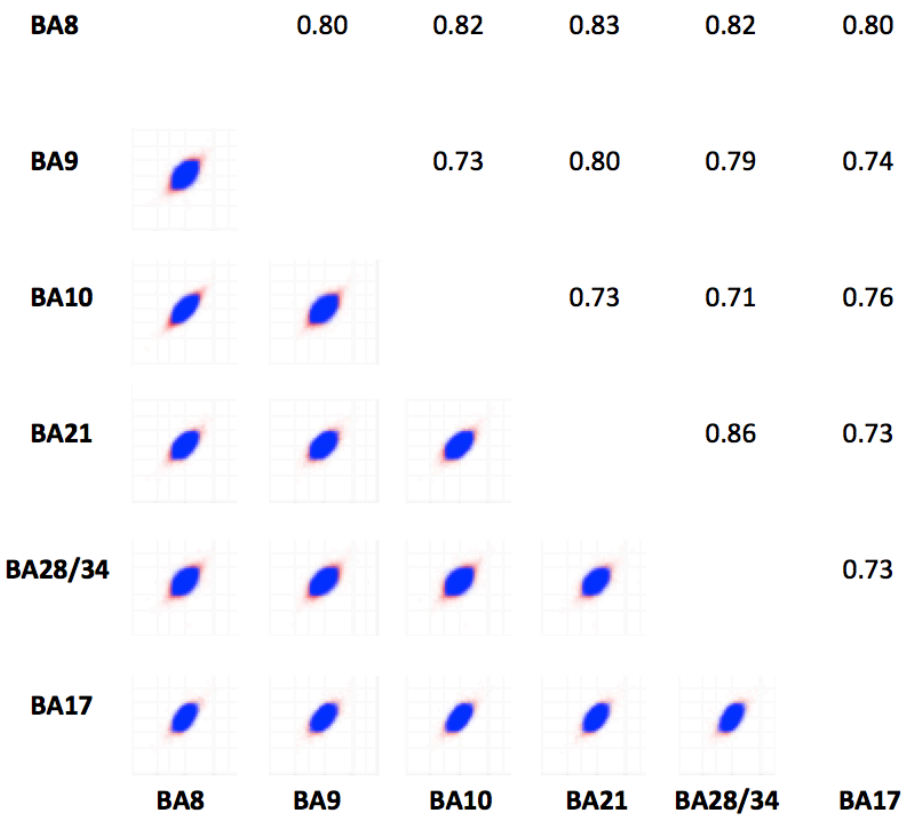


Figure 4-11. ASM scores are highly correlated across cortical regions. Lower triangular matrix: scatterplots of ASM scores, with probes characterized by absolute ASM scores ≥ 0.10 colored in red. Upper triangular matrix: correlations of ASM scores filtered for loci with at least one amplicon showing an absolute ASM score ≥ 0.10.

Supplementary Figure 5

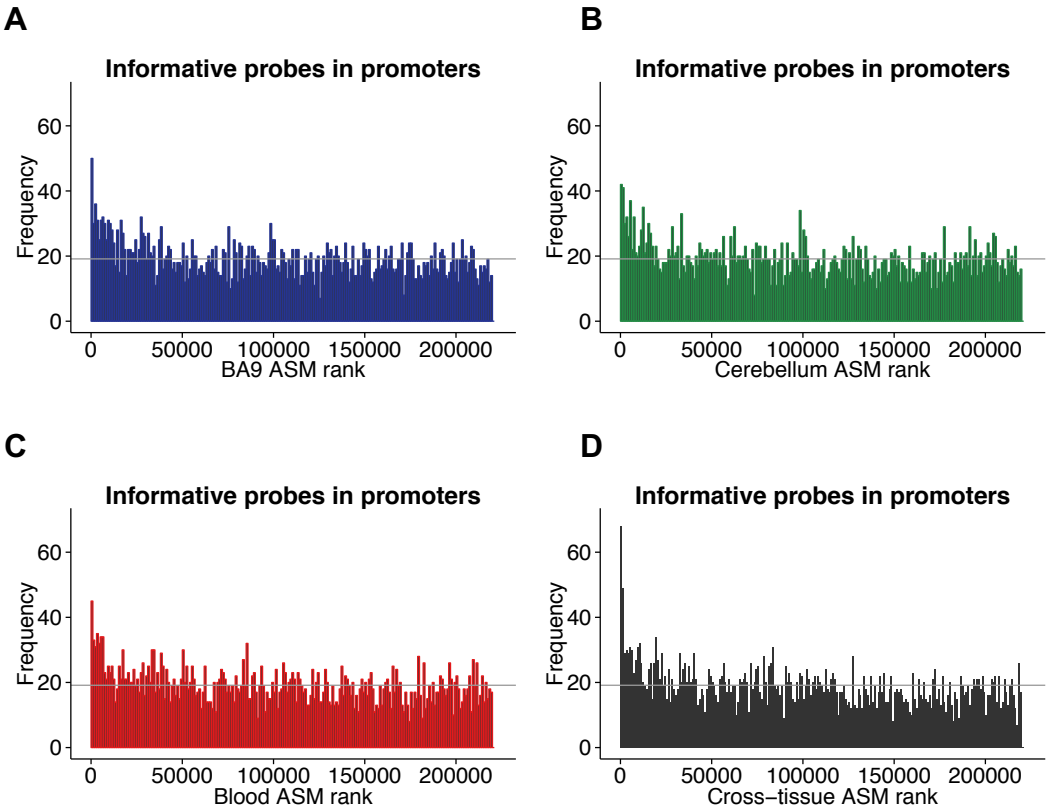
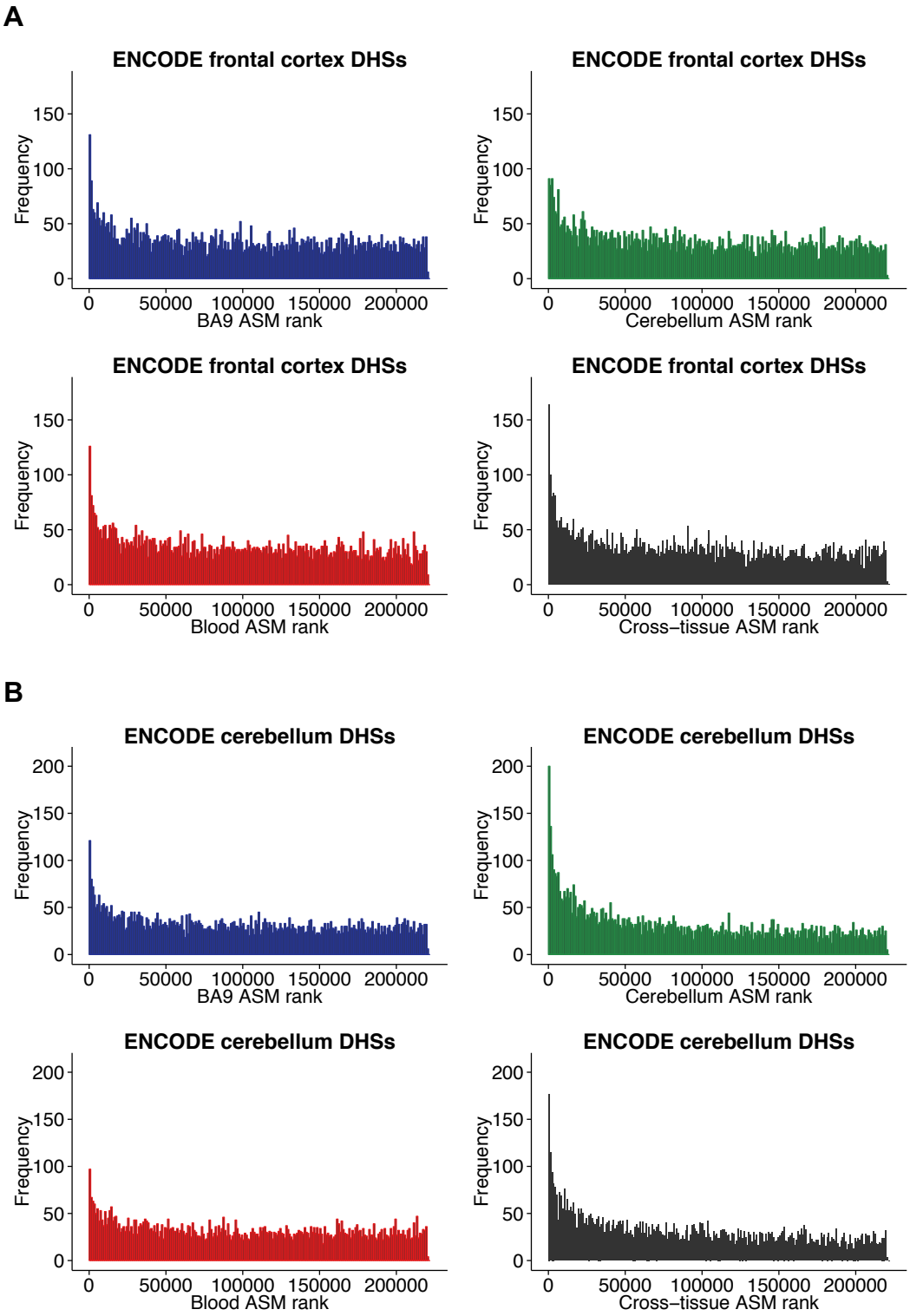


Figure 4-12. Informative probes annotated to genic promoter regions are enriched for elevated ASM scores. A Kruskal-Wallis non-parametric test showed a significantly different distribution of ranked ASM scores between coding, 5'UTR, intergenic, intronic, promoter and 3' UTR regions in informative cortex, cerebellum, blood and cross-tissue MSNP probes (*P* range = 2.21×10^{26} - 2.06×10^{12}). Post-hoc analysis showed that the differences were primarily driven by an enrichment of elevated ASM scores in annotated promoter regions for **(A)** cortex (BA9), **(B)** cerebellum, **(C)** whole blood and **(D)** cross-tissue ASM scores. The mean number of informative probes per bin is displayed as a grey horizontal line.

Supplementary Figure 6



C

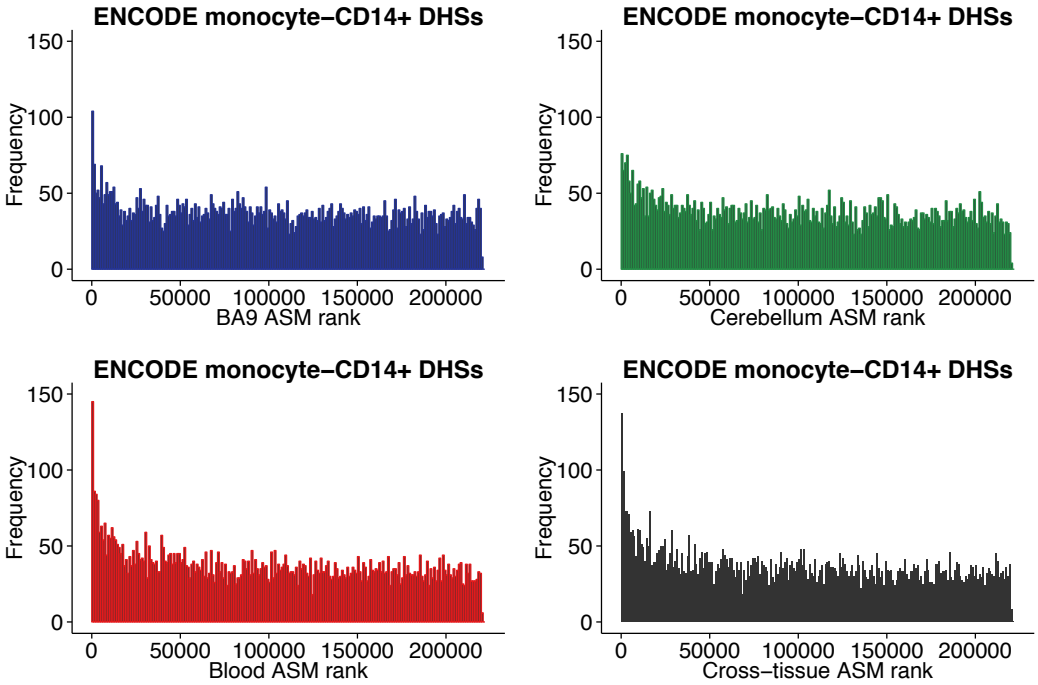


Figure 4-13. DNase I hypersensitive (DHS) sites are enriched for loci characterized by high ASM scores. DHS peaks overlapping informative MSNP probes are enriched for elevated ASM scores (see **Supplementary Table 3**). The histograms show the distribution of ASM score ranks for informative probes residing in ENCODE DHS peaks from frontal cortex (**A**), cerebellum (**B**) and monocyte-CD14+ cells (**C**). Shown is the distribution of ASM score ranks identified in cortex (BA9) (blue, top left), cerebellum (green, top right), whole blood (red, bottom left) and in the cross-tissue analysis (black, bottom right) across informative MSNP probes within ENCODE DHS peaks. All histograms show a clear peak in lower ranks with a uniform distribution across the rest of the rank spectrum, illustrating the higher than expected prevalence of probes with high ASM scores in the informative DHS regions.

Supplementary Figure 7

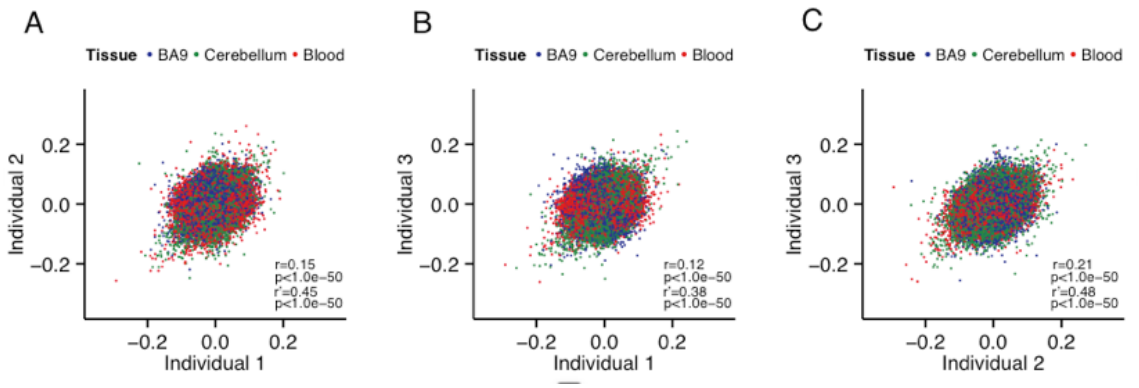
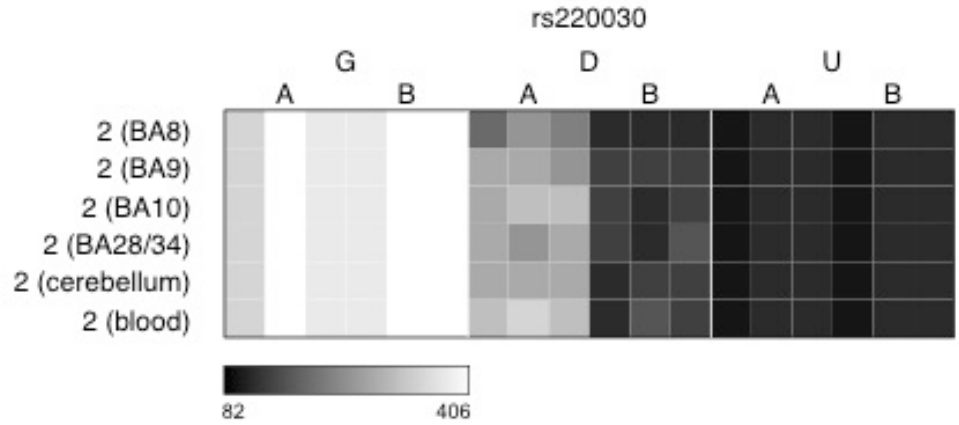


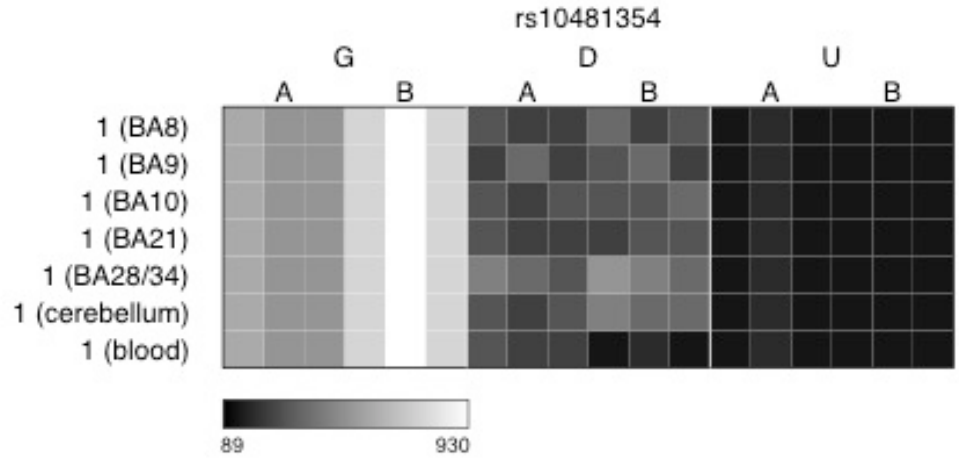
Figure 4-14. Individuals are more highly correlated for loci characterized by high ASM scores. ASM scores are significantly correlated between individuals, particularly at loci with an ASM score ≥ 0.10 for at least one of the two compared individuals (subset correlation r'). (**A**)-(**C**) Scatterplots of ASM scores across individuals plotted for each informative probe in cortex (BA9), cerebellum and whole blood.

Supplementary Figure 8

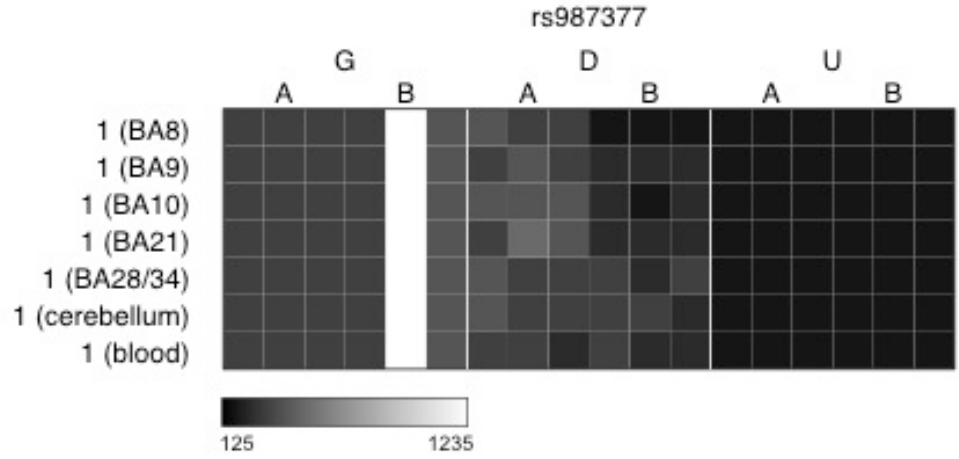
A



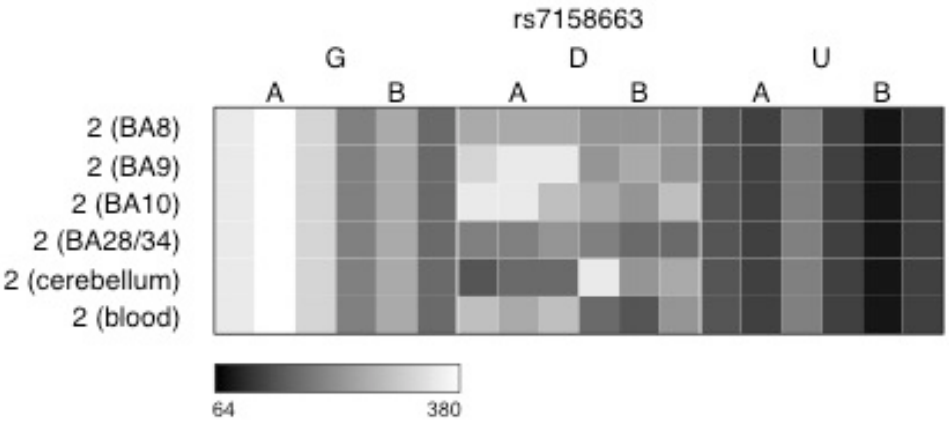
B



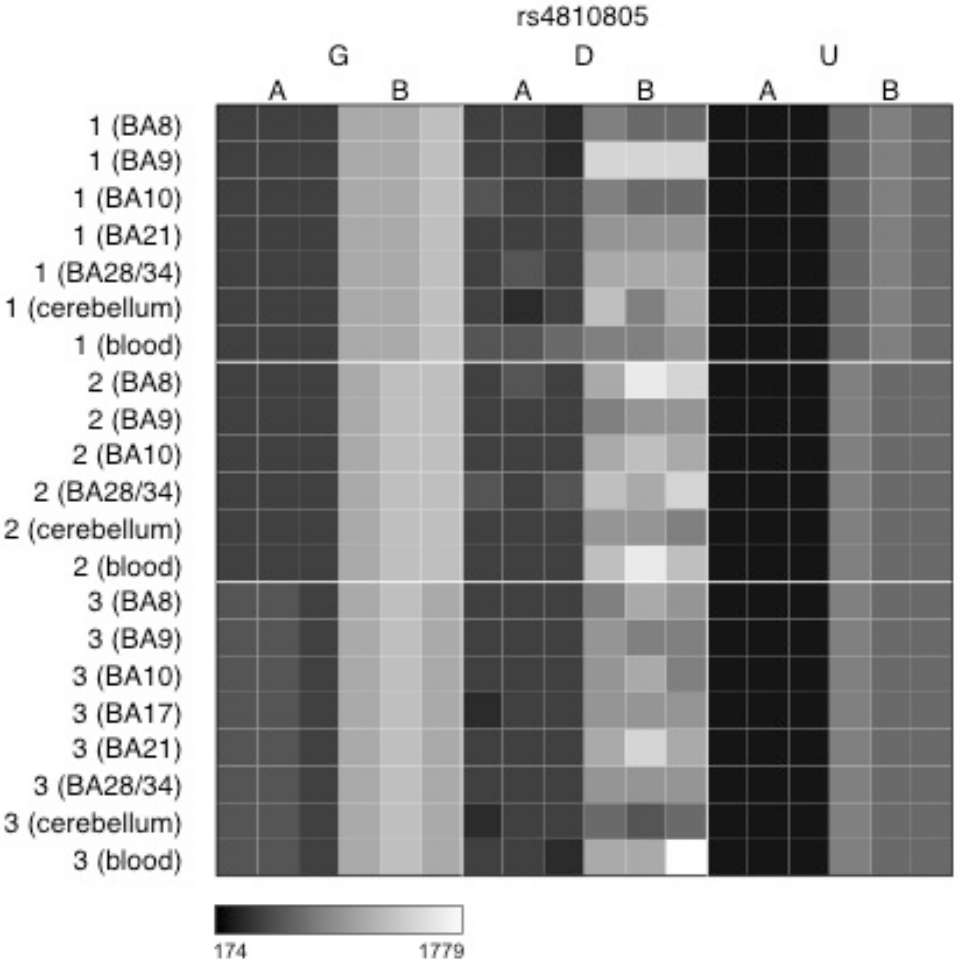
C



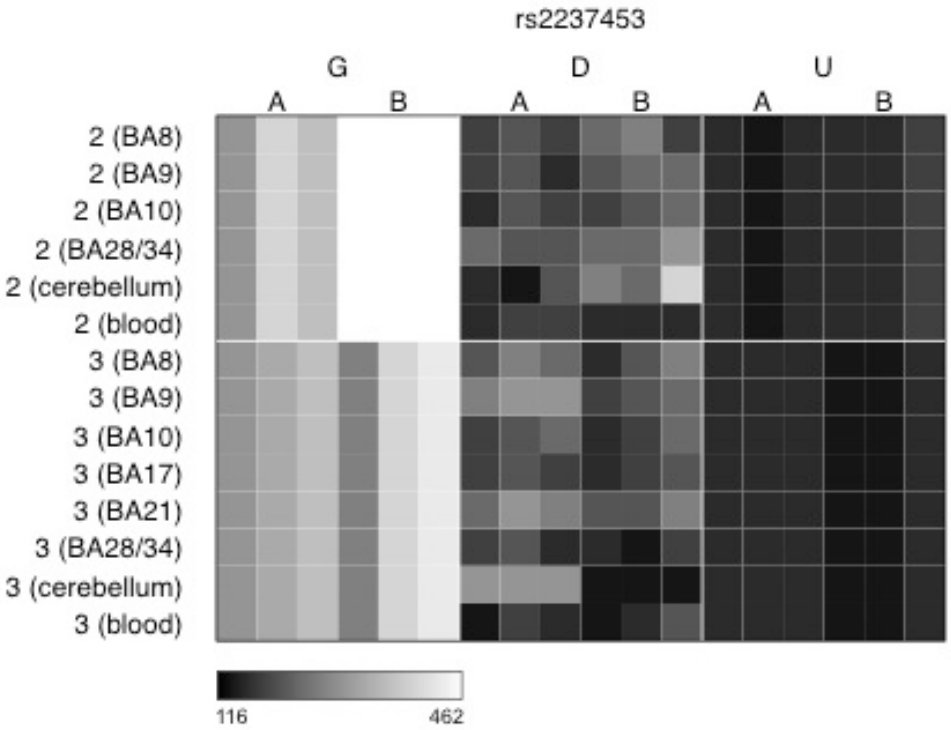
D


rs7158663
G
D
U
A
B
A
B
A
B
2 (BA8)
2 (BA9)
2 (BA10)
2 (BA28/34)
2 (cerebellum)
2 (blood)
64
380

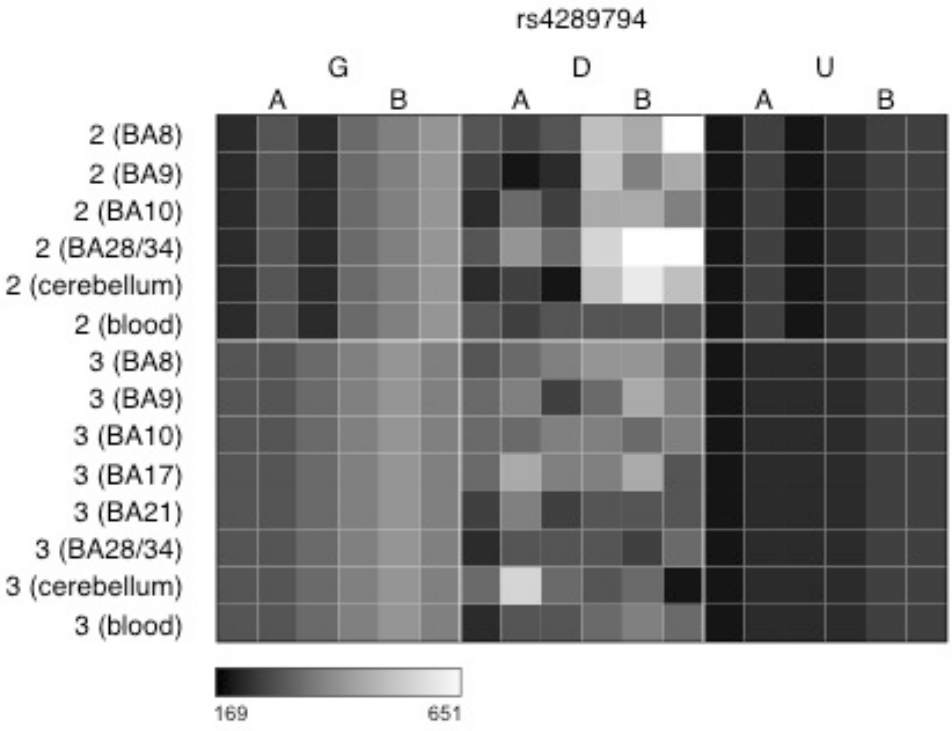
E


rs4810805
G
D
U
A
B
A
B
A
B
1 (BA8)
1 (BA9)
1 (BA10)
1 (BA21)
1 (BA28/34)
1 (cerebellum)
1 (blood)
2 (BA8)
2 (BA9)
2 (BA10)
2 (BA28/34)
2 (cerebellum)
2 (blood)
3 (BA8)
3 (BA9)
3 (BA10)
3 (BA17)
3 (BA21)
3 (BA28/34)
3 (cerebellum)
3 (blood)
174
1779

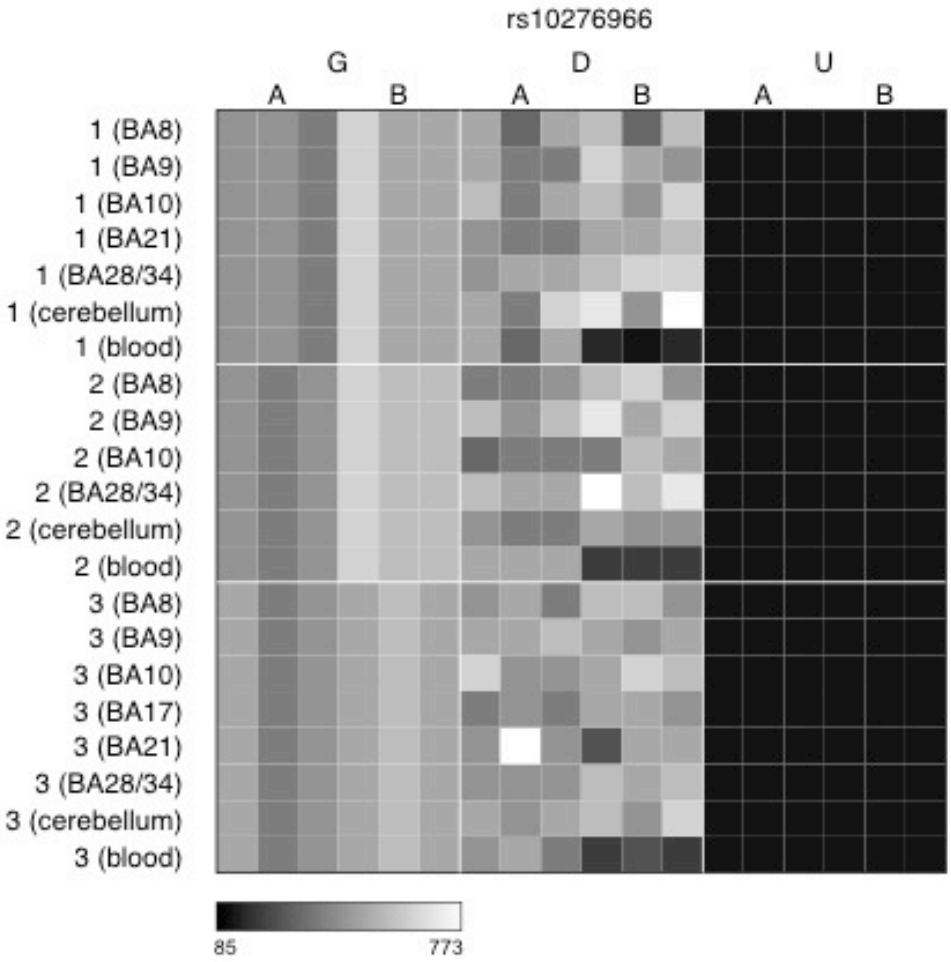
F


rs2237453
G
D
U
A
B
A
B
A
B
2 (BA8)
2 (BA9)
2 (BA10)
2 (BA28/34)
2 (cerebellum)
2 (blood)
3 (BA8)
3 (BA9)
3 (BA10)
3 (BA17)
3 (BA21)
3 (BA28/34)
3 (cerebellum)
3 (blood)
116
462

G


rs4289794
G
D
U
A
B
A
B
A
B
2 (BA8)
2 (BA9)
2 (BA10)
2 (BA28/34)
2 (cerebellum)
2 (blood)
3 (BA8)
3 (BA9)
3 (BA10)
3 (BA17)
3 (BA21)
3 (BA28/34)
3 (cerebellum)
3 (blood)
169
651

H


rs10276966
G
D
U
A
B
A
B
A
B
1 (BA8)
1 (BA9)
1 (BA10)
1 (BA21)
1 (BA28/34)
1 (cerebellum)
1 (blood)
2 (BA8)
2 (BA9)
2 (BA10)
2 (BA28/34)
2 (cerebellum)
2 (blood)
3 (BA8)
3 (BA9)
3 (BA10)
3 (BA17)
3 (BA21)
3 (BA28/34)
3 (cerebellum)
3 (blood)
85
773

I

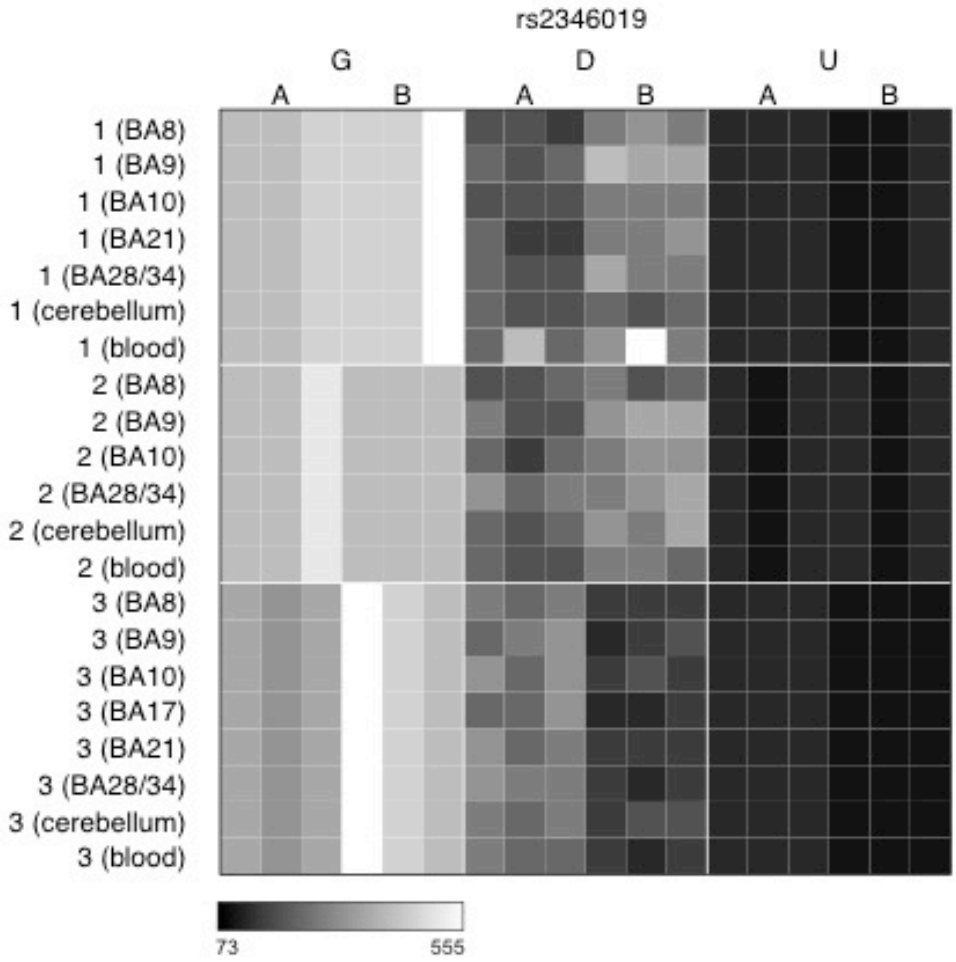


Figure 4-15. Figure 8. Several ASM sites were identified in known or suspected imprinted regions. Heatmaps display allele signal intensities for genomic DNA (G), MSRE-digested DNA (D) and fully unmethylated, MSRE-digested DNA (U) in all tissues. A and B denote the two alleles of the SNP and brightness represents the quantile normalized signal intensity, with the scale displayed below the heatmap. Shown are data from probes in the vicinity of known imprinted genes: (**A**) rs220030 (*SNRPN*), (**B**) rs10481354 (*DLGAP2*), (**C**) rs987377 (*AIM1*), (**D**) rs7158663 (*MEG3*), (**E**) rs4810805 (*BLCAP*) and (**F**) rs2237453 (*GRB10*). Also shown are data from probes in the vicinity of suspected imprinted loci: (**G**) rs4289794 (*TRAPPC9*), (**H**) rs10276966 (*EVX1*), (**I**) rs2346019 (*VTRNA2-1*).

Supplementary Figure 9

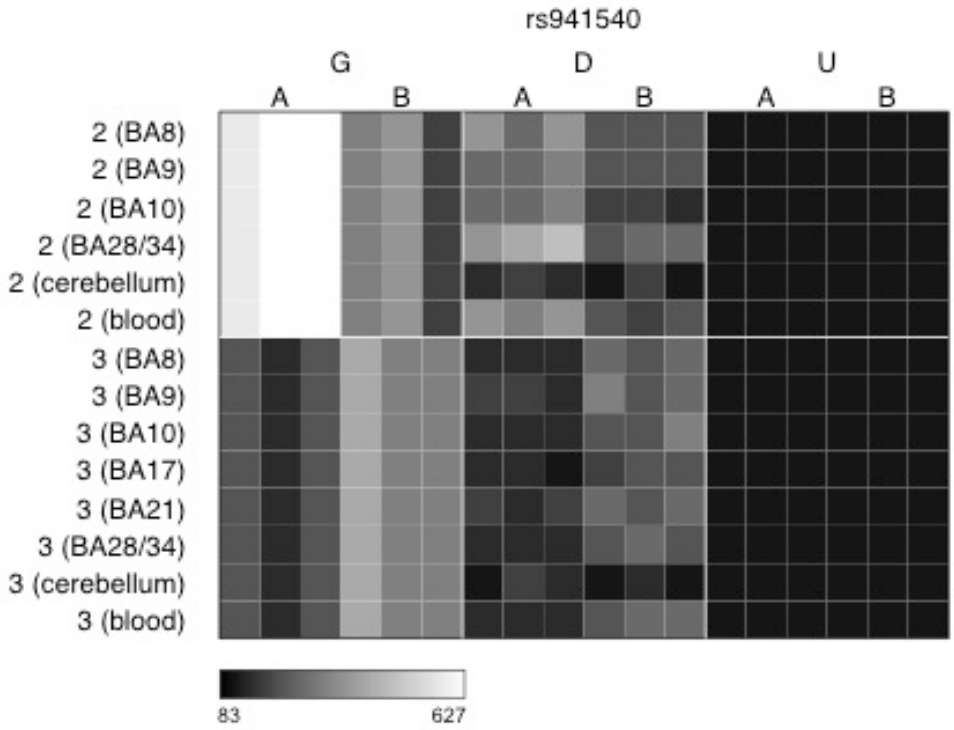


Figure 4-16. Some sites showing allelic-flipping in ASM are not located near known or suspected imprinted regions. The heatmap displays allele signal intensities for genomic DNA (G), MSRE-digested DNA (D) and fully unmethylated, MSRE-digested DNA (U) in all tissues. A and B denote the two alleles of the SNP and brightness represents the quantile normalized signal intensity, with the scale displayed below the heatmap. The heatmap shows rs941540 (*ITPKI*), which is located in a region not previously shown to be genomically imprinted.

Supplementary Table 1

Table 4-8. Average absolute ASM scores for each sample.

	BA8	BA9	BA10	BA17	BA21	BA28/34	Cerebellum	Blood	Average
Individual 1	0.026	0.025	0.028		0.023	0.027	0.023	0.028	0.026
Individual 2	0.023	0.027	0.024			0.026	0.028	0.026	0.025
Individual 3	0.025	0.027	0.025	0.024	0.023	0.023	0.030	0.024	0.025
Average	0.025	0.026	0.026	0.024	0.023	0.025	0.027	0.026	0.025

Supplementary Table 2

Table 4-9. Percentage of amplicons characterized by an ASM score ≥ 0.10 for each sample.

	BA8	BA9	BA10	BA17	BA21	BA28/34	Cerebellum	Blood	Average
Individual 1	0.68%	0.40%	0.84%		0.37%	0.80%	0.56%	0.97%	0.66%
Individual 2	0.42%	0.64%	0.44%			0.64%	1.36%	1.03%	0.76%
Individual 3	0.47%	0.73%	0.55%	0.43%	0.31%	0.34%	1.50%	0.51%	0.60%
Average	0.52%	0.59%	0.61%	0.43%	0.34%	0.59%	1.14%	0.84%	0.63%

Supplementary Table 3

Table 4-10. Enrichment of high ASM scores in DNase I hypersensitivity peaks spanning informative MSNP probes.

Tissue/cell dataset from ENCODE	Number of informative probes in DHS peaks	Enrichment *P* value*			
		BA9 ASM score	Cerebellum ASM score	Blood ASM score	Cross-tissue ASM score
Frontal cortex	7746	4.59×10^{-38}	5.69×10^{-58}	1.41×10^{-43}	1.19×10^{-81}
Cerebrum frontal	8294	8.84×10^{-51}	1.75×10^{-80}	3.19×10^{-42}	2.92×10^{-107}
Cerebellum	7228	2.10×10^{-47}	3.51×10^{-220}	1.09×10^{-27}	4.74×10^{-181}
Monocytes - CD14+	8316	4.98×10^{-15}	4.13×10^{-24}	7.41×10^{-52}	5.43×10^{-53}
Naïve B cell	3908	4.71×10^{-31}	2.10×10^{-37}	2.99×10^{-96}	1.35×10^{-99}
Heart	7268	5.05×10^{-42}	1.19×10^{-54}	1.78×10^{-55}	3.86×10^{-93}
Fibroblast	4316	7.50×10^{-25}	1.67×10^{-24}	1.85×10^{-20}	2.07×10^{-42}
H1-hesc	7208	2.92×10^{-44}	8.77×10^{-67}	3.52×10^{-56}	1.98×10^{-93}

* *P* values are based on a Wilcoxon rank-sum test comparing ASM ranks in DHS and non-DHS informative probes

Supplementary Table 4

Table 4-11. Correlations of ASM scores between individuals.

	Correlation (*r*)				Subset correlation (*r'*)*			
	BA9	**Cerebellum**	**Blood**	**All tissues+**	**BA9**	**Cerebellum**	**Blood**	**All tissues+**
Individual 1 vs. 2	0.15	0.19	0.12	0.15	0.45	0.51	0.41	0.45
Individual 1 vs. 3	-0.01	0.23	0.14	0.12	0.12	0.55	0.41	0.38
Individual 2 vs. 3	0.15	0.28	0.17	0.21	0.34	0.52	0.53	0.48

* The subset correlation *r'* is calculated using only probes for which at least one of the two compared individuals exhibit an ASM score ≥ 0.10 in the given tissue.

+ Cerebellum, cortex (BA9) and whole blood

Supplementary Table 5

Table 4-12. Top 20 variable ASM sites in whole blood, defined by range of ASM scores.

Rank	SNP ID	Location	Associated gene(s)	Blood ASM scores			Range per tissue		
				Ind 1	Ind 2	Ind 3	Blood	Cerebellum	BA9
1	SNP_A-8579417 (rs1538116)	1q24.1	*MGST3, LOC400794*	-	-0.29	0.06	0.35	0.24	0.32
2	SNP_A-8312502 (rs17469830)	14q12	*ARHGAP5, NUBPL, C14orf128*	-0.07	0.21	-	0.28	0.12	0.03
3	SNP_A-1802351 (rs509062)	20q13.32	*PPP4R1L*	-0.16	-	0.08	0.25	0.13	0.05
4	SNP_A-2019421 (rs2244352)	21q22.2	*WRB*	-0.19	0.05	-	0.24	0.36	0.29
5	SNP_A-8324660 (rs6438837)	3q21.2	*KALRN*	-0.14	0.09	-0.06	0.24	0.16	0.10
6	SNP_A-4232094 (rs8179356)	1p21.2	*FRRS1*	-	0.08	-0.16	0.23	0.11	0.07
7	SNP_A-1855770 (rs3764124)	13q34	*CUL4A*	-0.11	0.12	0.06	0.23	0.16	0.15
8	SNP_A-4219174 (rs2346019)	5q31.1	*TGFBI, VTRNA2*[b]	0.04	0.07	-0.16	0.23	0.23	0.30
9	SNP_A-4208914 (rs927651)	20q13.2	*CYP24A1*	0.15	-0.08	0.00	0.23	0.25	0.17
10	SNP_A-2275375 (rs7694398)	4p16.1	*CNO, KIAA0232*	0.12	-0.10	-0.01	0.22	0.08	0.05

11	SNP_A-1910565 (rs1040762)	20p11.23	*RIN2, SLC24A3*	0.01	-0.20		0.21	0.04	0.09
12	SNP_A-8448479 (rs12837024)	Xq22.3	*CLDN2, RIPPLY1*	-0.06	-	0.15	0.21	0.10	0.15
13	SNP_A-2074691 (rs7065328)	Xq13.1	*TEX11*	-0.10	-	0.10	0.21	0.08	0.01
14	SNP_A-8330843 (rs896040)	18q23	*GALR1, MBP*	0.14	-0.05	-0.06	0.21	0.11	0.06
15	SNP_A-8605034 (rs17799404)	4q13.2	*CENPC1, LOC100144602*	0.13	-	-0.07	0.20	0.05	0.01
16	SNP_A-8633376 (rs318657)	5p15.33	*LOC285577, C5orf38*	-0.11	0.09	-	0.20	0.11	0.08
17	SNP_A-4239614 (rs1890070)	1q41	*HHIPL2, DUSP10*	-0.16	0.04	-	0.20	0.02	0.02
18	SNP_A-2256850 (rs1164193)	3q13.13	*PVRL3, FLJ25363*	0.08	-0.12	-	0.20	0.01	0.08
19	SNP_A-1869018 (rs1984713)	4p12	*CNGA1*	-	-0.18	0.02	0.20	0.15	0.14
20	SNP_A-8436908 (rs4810805)	20q11.23	*CTNNBL1, BLCAP*[a]	-0.14	0.05	0.06	0.20	0.05	0.07

[a] Known imprinted gene
[b] Suspected imprinted gene

Supplementary Table 6

Table 4-13. Top 20 variable ASM sites in cerebellum, defined by range of ASM scores.

Rank	SNP ID	Location	Associated gene(s)	Cerebellum ASM scores				Range per tissue	
				Ind 1	Ind 2	Ind 3	Cerebellum	Blood	BA9
1	SNP_A-2019421 (rs2244352)	21q22.2	*WRB*	-0.22	0.14	-	0.36	0.24	0.29
2	SNP_A-8492800 (rs2237453)	7p12.1	*GRB10*[a]	-	0.06	-0.23	0.30	0.11	0.08
3	SNP_A-8291694 (rs941540)	14q32.12	*ITPK1*	-	0.10	-0.17	0.27	0.07	0.13
4	SNP_A-8717059 (rs12713666)	2p13.3	*ARHGAP25*	-0.08	0.18	0.04	0.26	0.15	0.21
5	SNP_A-4208914 (rs927651)	20q13.2	*CYP24A1*	0.08	-0.17	-0.06	0.25	0.23	0.17
6	SNP_A-8596407 (rs10778439)	12q23.3	*NUAK1, C12orf75*	0.00	-	-0.25	0.25	0.03	0.09
7	SNP_A-1788232 (rs11854691)	15q11.2	*C15orf2*[a], *PWRN2, NDN*[a]	-0.15	0.10	-	0.25	0.06	0.17
8	SNP_A-2118217 (rs1695824)	1p36.33	*VWA1, TMEM88B*	-0.22	-	0.03	0.25	0.16	0.13
9	SNP_A-2252209 (rs940596)	15q11.2	*C15orf2*[a], *PWRN2, NDN*[a]	-0.10	0.11	-0.13	0.25	0.04	0.09
10	SNP_A-8713358 (rs16890883)	4p15.33	*CPEB2, LOC152742*	0.04	-0.21	0.01	0.24	0.14	0.17

11	SNP_A-8397290 (rs4289794)	8q24.3	*TRAPPC9*[b]	-	0.10	-0.14	0.24	0.09	0.08
12	SNP_A-8579417 (rs1538116)	1q24.1	*MGST3, LOC400794*	-	-0.24	0.00	0.24	0.35	0.32
13	SNP_A-8321422 (rs3759917)	15q26.1	*ST8SIA2, SLCO3A1*	0.00	0.24	-	0.24	0.12	0.14
14	SNP_A-1871907 (rs9460408)	6p22.3	*ID4, MIR548A1*	-	0.17	-0.06	0.23	0.12	0.08
15	SNP_A-4219174 (rs2346019)	5q31.1	*TGFB, VTRNA2*	0.00	0.10	-0.13	0.23	0.23	0.30
16	SNP_A-8353238 (rs9570315)	13q21.2	*MIR3169, TDRD3*	-0.07	-	0.17	0.23	0.16	0.09
17	SNP_A-1880775 (rs6116750)	20p12.3	*PROKR2, LOC643406*	-	0.12	-0.11	0.23	0.13	0.18
18	SNP_A-2244974 (rs941483)	14q32.12	*ASB2*	0.01	-0.17	0.05	0.23	0.04	0.05
19	SNP_A-8638465 (rs11633486)	15q11.2	*C15orf2*[a], *PWRN2, NDN*[a]	-0.13	0.10	-	0.23	0.11	0.16
20	SNP_A-8437947 (rs2858602)	22q13.33	*C22orf34, FAM19A5*	-0.07	0.11	0.16	0.22	0.06	0.15

[a] Known imprinted gene
[b] Suspected imprinted gene

Supplementary Table 7

Table 4-14. Top 20 variable ASM sites in cortex (BA9), as defined by range of ASM scores.

Rank	SNP ID	Location	Associated gene(s)	BA9 ASM scores			Range per tissue		
				Ind 1	Ind 2	Ind 3	BA9	Cerebellum	Blood
1	SNP_A-8579417 (rs1538116)	1q24.1	*MGST3, LOC400794*	-	-0.24	0.08	0.32	0.24	0.35
2	SNP_A-8424056 (rs3922835)	18q12.1	*CDH2*[b], *CHST9*	-0.15	-	0.17	0.32	0.18	0.04
3	SNP_A-4219174 (rs2346019)	5q31.1	*TGFBI, VTRNA2*	0.10	0.13	-0.17	0.30	0.23	0.23
4	SNP_A-2019421 (rs2244352)	21q22.2	*WRB*	-0.15	0.14	-	0.29	0.36	0.24
5	SNP_A-8448997 (rs17346747)	8p23.2	*CSMD1*	-0.11	0.01	0.15	0.26	0.10	0.03
6	SNP_A-8692937 (rs4605656)	4q24	*CXXC4, TACR3*	-0.19	-	0.06	0.25	0.16	0.18
7	SNP_A-8658815 (rs9807659)	18q11.2	*LAMA3*	-0.08	0.02	0.17	0.24	0.12	0.10
8	SNP_A-8311269 (rs403594)	1p32.3	*ACOT11*	-0.12	0.03	0.12	0.24	0.11	0.02
9	SNP_A-8634251 (rs1209228)	14q24.2	*RGS6, SIPA1L1*	-0.18	0.06	0.01	0.24	0.19	0.15

10	SNP_A-1951087 (rs9357087)	6p24.2	*NEDD9*	-	0.03	-0.19	0.23	0.07	0.07
11	SNP_A-8487535 (rs503755)	2q21.3	*TMEM163*	-	0.10	-0.12	0.22	0.01	0.09
12	SNP_A-8329713 (rs16825906)	3q13.31	*LSAMP, IGSF11, LOC285194*	-0.04	0.04	0.19	0.22	0.21	0.13
13	SNP_A-8301613 (rs4380126)	18q23	*C18orf62, TSHZ1*	0.10	-0.02	-0.12	0.22	0.17	0.10
14	SNP_A-8368457 (rs7307323)	12p11.22	*FAR2, CCDC91*	-	0.19	-0.03	0.21	0.09	0.05
15	SNP_A-8399845 (rs197493)	6q24.2	*HIVEP2, LOC153910*	-0.08	-0.02	0.13	0.21	0.15	0.03
16	SNP_A-8717059 (rs12713666)	2p13.3	*ARHGAP25*	-0.13	0.08	0.02	0.21	0.26	0.15
17	SNP_A-8515165 (rs7892015)	Xq21.1	*POU3F4, SH3BGRL*	-0.03	-	0.18	0.21	0.15	0.10
18	SNP_A-8613922 (rs1009800)	13q21.33	*KLHL1, PCDH9*	-0.06	0.03	0.16	0.21	0.11	0.05
19	SNP_A-8482424 (rs10880337)	12q12	*ADAMTS20, PRICKLE1*	-0.11	0.10	0.05	0.21	0.15	0.08
20	SNP_A-2246435 (rs7070457)	10q21.1	*ZWINT*	0.12	-	-0.09	0.20	0.19	0.13

[b] Suspected imprinted gene

Supplementary Table 8

Table 4-15. Illumina 450K probes within 1kb of 100 top-ranked ASM sites.

Top 100 ASM sites	SNPs	CpG sites
Blood ASM	27	102
Cerebellum ASM	20	87
BA9 ASM	40	143
Cross-tissue ASM	34	148
Tissue-specific ASM	22	50
Variable ASM	22	65
Variable ASM (blood)	19	58
Variable ASM (cerebellum)	18	59
Variable ASM (BA9)	20	56

Supplementary Table 9

Table 4-16. Illumina 450K probes within 1kb of the top 100 ASM SNPs in whole blood.

Rank	ASM in blood	SNP	Chr	Mapinfo SNP	CpG site	Mapinfo CpG	Distance
3	0.23	rs10234308	7	22528590	cg10016647	22528033	557
4	0.23	rs927000	20	43598705	cg17104026	43598376	329
10	0.22	rs4687210	3	191046770	cg25342125	191046380	390
					cg19066520	191046547	223
					cg23528751	191046721	49
					cg00473167	191046872	102
					cg17596905	191046949	179
					cg23817170	191046997	227
					cg09688588	191047460	690
					cg14972328	191046866	96
					cg26325444	191047058	288
12	0.22	rs12978286	19	17861868	cg22317846	17862385	517
					cg02639359	17862017	149
					cg26307105	17861017	851
					cg08206267	17862104	236
					cg11737334	17862210	342
					cg24783510	17862150	282
13	0.22	rs10481354	8	1704030	cg19131313	1704013	17
					cg17159058	1704075	45
14	0.21	rs1542180	2	177028767	cg12634591	177028606	161
					cg06145336	177027931	836
					cg10304824	177028804	37
					cg13316854	177027856	911
					cg09387749	177028680	87
					cg25938806	177028162	605
					cg01128482	177028621	146
					cg00005847	177029073	306
					cg24541426	177029459	692
					cg23981871	177028714	53
					cg05021643	177029608	841
25	0.20	rs17312325	X	68442113	cg00427681	68441389	724
					cg03350825	68442327	214
					cg22337892	68441690	423
27	0.20	rs1024611	17	32579788	cg13648045	32579022	766
30	0.20	rs12493005	3	176912899	cg03356026	176912588	311
37	0.19	rs12510921	4	185189278	cg04960065	185189300	22
43	0.19	rs906805	2	28604879	cg01273330	28605224	345
44	0.18	rs220030	15	25199768	cg21746532	25198793	975
					cg26875073	25200490	722
					cg02152271	25199270	498
					cg02125271	25200406	638
					cg03858387	25199164	604
					cg01432432	25199028	740
					cg08372135	25199057	711
					cg18506672	25200253	485
					cg12298755	25199713	55

48	0.18	rs2252397	7	140225750	cg01499591	140226367	617
50	0.18	rs999556	5	150473674	cg23290159	150474234	560
56	0.18	rs9526072	13	32883583	cg26941801	32882845	738
					cg11448807	32884114	531
57	0.18	rs7010076	8	105342827	cg13045555	105342365	462
					cg04554929	105342491	336
					cg15172053	105342214	613
					cg23108931	105342351	476
					cg17696517	105343133	306
58	0.18	rs485219	8	70065359	cg05959464	70064456	903
62	0.18	rs7210715	17	7196390	cg00901493	7197375	985
64	0.17	rs17366322	2	11272784	cg05517905	11273075	291
					cg09045196	11272424	360
					cg12453288	11273229	445
					cg10334385	11272868	84
					cg03251967	11273291	507
					cg15704155	11272399	385
					cg27641709	11272899	115
67	0.17	rs16942894	12	114026006	cg12492938	114025383	623
70	0.17	rs9314433	8	1467648	cg00598912	1468625	977
71	0.17	rs2071780	X	47341821	cg26176649	47342033	212
					cg02124059	47341898	77
					cg22238863	47342101	280
					cg22713892	47342160	339
					cg23534593	47342469	648
					cg21161328	47342498	677
					cg24559073	47342471	650
					cg19867709	47342478	657
					cg04302192	47341740	81
					cg04575501	47342663	842
					cg14388993	47342690	869
					cg01887353	47342480	659
78	0.17	rs2332036	3	121714391	cg20356878	121714668	277
					cg11854227	121714109	282
83	0.17	rs4810485	20	44747947	cg21601405	44747006	941
					cg16686951	44747351	596
86	0.16	rs10883857	10	105127587	cg23418095	105128462	875
					cg14125903	105126693	894
					cg00245789	105128183	596
					cg20188621	105127192	395
					cg11667101	105127581	6
					cg10301212	105126702	885
					cg26978691	105128013	426
					cg18732587	105127670	83
					cg25713684	105127687	100
					cg18586343	105127615	28
					cg20684973	105127632	45
					cg16500454	105127701	114
					cg01360325	105127811	224
88	0.16	rs4075600	16	86929340	cg07824497	86930033	693
					cg08569253	86930198	858

					cg05490489	86930336	996
90	0.16	rs10411704	19	35800662	cg22238209	35800743	81
					cg07918620	35800925	263
					cg15761414	35801014	352
					cg02776658	35800589	73

Supplementary Table 10

Table 4-17. Illumina 450K probes within 1kb of the top 100 ASM SNPs in cerebellum.

Rank	Cerebellum ASM score	SNP	Chr	Mapinfo SNP	CpG site	Mapinfo CpG	Distance
2	0.29	rs1003533	5	131755651	cg03347934	131754928	723
3	0.26	rs7959070	12	92805682	cg25400182	92806045	363
4	0.25	rs10234308	7	22528590	cg10016647	22528033	557
6	0.24	rs12493005	3	176912899	cg03356026	176912588	311
7	0.24	rs1542180	2	177028767	cg13316854	177027856	911
					cg06145336	177027931	836
					cg25938806	177028162	605
					cg12634591	177028606	161
					cg01128482	177028621	146
					cg09387749	177028680	87
					cg23981871	177028714	53
					cg10304824	177028804	37
					cg00005847	177029073	306
					cg24541426	177029459	692
					cg05021643	177029608	841
9	0.24	rs3098382	5	71462632	cg20954533	71462729	97
12	0.23	rs7158663	14	101319424	cg19509303	101318703	721
					cg07460524	101319526	102
21	0.21	rs12413873	10	71212091	cg13358349	71211210	881
					cg12712771	71211212	879
					cg26465120	71211268	823
					cg11279933	71211301	790
					cg01566955	71211468	623
					cg18382422	71211551	540
					cg12610070	71211762	329
					cg14388049	71211838	253
24	0.21	rs2546890	5	158759900	cg07622001	158758903	997
45	0.19	rs10861495	12	106142981	cg23539745	106142108	873
63	0.18	rs971649	8	124732275	cg13177458	124732088	187
65	0.18	rs7247675	19	28285519	cg17155612	28284698	821
					cg02942594	28284741	778
					cg03170611	28284813	706
					cg22473312	28284823	696
					cg08220598	28284952	567
					cg17871739	28284973	546
					cg24415565	28284988	531
					cg25246092	28285128	391
					cg03751272	28285308	211
					cg06949053	28285395	124
72	0.18	rs3762199	20	18487792	cg26234223	18487771	21
					cg01866064	18487951	159

					cg27634549	18487976	184
					cg23490046	18487999	207
					cg15613982	18488017	225
					cg16156954	18488094	302
					cg07973146	18488185	393
					cg02081051	18488312	520
					cg22243844	18488366	574
					cg13313269	18488622	830
73	0.18	rs220030	15	25199768	cg21746532	25198793	975
					cg01432432	25199028	740
					cg08372135	25199057	711
					cg03858387	25199164	604
					cg02152271	25199270	498
					cg12298755	25199713	55
					cg18506672	25200253	485
					cg02125271	25200406	638
					cg26875073	25200490	722
76	0.18	rs2244352	21	40757973	cg01322405	40757317	656
					cg00606841	40757691	282
					cg22858667	40757750	223
					cg26710963	40757899	74
					cg09916765	40758208	235
					cg21297395	40758325	352
					cg21459460	40758346	373
84	0.18	rs9469529	6	33587443	cg08355863	33587496	53
					cg24603235	33588219	776
					cg19889152	33588278	835
					cg14639225	33588302	859
85	0.18	rs1342331	6	144066159	cg06759993	144065203	956
					cg08482167	144066847	688
86	0.18	rs7185244	16	86546887	cg08142918	86546374	513
					cg03697918	86546628	259
					cg24908603	86546631	256
					cg07056644	86546785	102
					cg27453745	86546938	51
					cg09338251	86546979	92
					cg06834912	86547203	316
					cg04787888	86547322	435
					cg01243371	86547386	499
					cg00551679	86547530	643
					cg02783918	86547544	657
87	0.18	rs12582221	12	125323240	cg17534092	125322578	662
99	0.17	rs6713344	2	112814045	cg09071112	112813174	871
					cg06723287	112813613	432
					cg11931949	112814041	4
					cg19917083	112814452	407

Supplementary Table 11

Table 4-18. Illumina 450K probes sites within 1kb of the top 100 ASM SNPs in BA9.

Rank	BA9 ASM score	SNP	Chr	Mapinfo SNP	CpG site	Mapinfo CpG	Distance
1	0.24	rs12493005	3	176912899	cg03356026	176912588	311
3	0.23	rs10234308	7	22528590	cg10016647	22528033	557
7	0.19	rs3098382	5	71462632	cg20954533	71462729	97
9	0.19	rs3121125	1	145382124	cg23685870	145382340	216
					cg07117663	145382963	839
10	0.19	rs1542180	2	177028767	cg13316854	177027856	911
					cg06145336	177027931	836
					cg25938806	177028162	605
					cg12634591	177028606	161
					cg01128482	177028621	146
					cg09387749	177028680	87
					cg23981871	177028714	53
					cg10304824	177028804	37
					cg00005847	177029073	306
					cg24541426	177029459	692
					cg05021643	177029608	841
14	0.18	rs17303015	5	43194137	cg10641714	43193681	456
15	0.18	rs4871852	8	23020396	cg23051664	23020966	570
					cg10964421	23021093	697
					cg01456571	23021325	929
20	0.18	rs6874600	5	73945738	cg20658167	73944869	869
25	0.18	rs7073504	10	60271987	cg07567376	60271505	482
					cg09696939	60272079	92
					cg07857251	60272754	767
					cg19940537	60272938	951
29	0.17	rs6929846	6	26458265	cg03553910	26457928	337
					cg00369351	26458121	144
					cg27318050	26458189	76
					cg22814929	26458213	52
					cg01542644	26458326	61
					cg11239695	26458489	224
30	0.17	rs971649	8	124732275	cg13177458	124732088	187
33	0.17	rs1139405	17	79478019	cg12793711	79477853	166
					cg26857837	79477974	45
					cg06713261	79478173	154
					cg15906692	79478715	696
					cg17317962	79478839	820
37	0.17	rs4687210	3	191046770	cg25342125	191046380	390
					cg19066520	191046547	223
					cg23528751	191046721	49
					cg14972328	191046866	96
					cg00473167	191046872	102

					cg17596905	191046949	179
					cg23817170	191046997	227
					cg26325444	191047058	288
					cg09688588	191047460	690
38	0.17	rs1774846	1	202994374	cg03639170	202993923	451
					cg00501542	202995246	872
39	0.17	rs12629627	3	73629548	cg14228812	73630106	558
41	0.17	rs6466131	7	106408569	cg09735627	106407970	599
42	0.17	rs1054497	4	174089048	cg27433088	174089019	29
					cg10817093	174089402	354
					cg06965122	174089515	467
					cg05897918	174089748	700
					cg18029791	174089755	707
					cg08422745	174089978	930
43	0.17	rs17366322	2	11272784	cg15704155	11272399	385
					cg09045196	11272424	360
					cg10334385	11272868	84
					cg27641709	11272899	115
					cg05517905	11273075	291
					cg12453288	11273229	445
					cg03251967	11273291	507
44	0.17	rs7305595	12	3562234	cg23794848	3563020	786
45	0.16	rs3762199	20	18487792	cg26234223	18487771	21
					cg01866064	18487951	159
					cg27634549	18487976	184
					cg23490046	18487999	207
					cg15613982	18488017	225
					cg16156954	18488094	302
					cg07973146	18488185	393
					cg02081051	18488312	520
					cg22243844	18488366	574
					cg13313269	18488622	830
51	0.16	rs10883857	10	105127587	cg14125903	105126693	894
					cg10301212	105126702	885
					cg20188621	105127192	395
					cg11667101	105127581	6
					cg18586343	105127615	28
					cg20684973	105127632	45
					cg18732587	105127670	83
					cg25713684	105127687	100
					cg16500454	105127701	114
					cg01360325	105127811	224
					cg26978691	105128013	426
					cg00245789	105128183	596
					cg23418095	105128462	875
52	0.16	rs1237543	13	24739364	cg03903831	24739249	115
57	0.16	rs1673087	19	37264190	cg05616754	37263860	330

					cg03430717	37263925	265
					cg27080211	37263997	193
59	0.16	rs6713344	2	112814045	cg09071112	112813174	871
					cg06723287	112813613	432
					cg11931949	112814041	4
					cg19917083	112814452	407
62	0.16	rs266388	15	67239366	cg09482748	67239225	141
64	0.16	rs9331942	8	27455114	cg01572979	27455004	110
65	0.15	rs2735971	11	2021649	cg27372170	2021103	546
					cg16574793	2022324	675
					cg09452478	2022386	737
72	0.15	rs11089197	22	18121334	cg16048163	18121127	207
					cg03887534	18121194	140
					cg24160660	18121281	53
					cg19898043	18121309	25
					cg23077364	18121349	15
					cg10625092	18121415	81
					cg24756391	18121424	90
					cg16593229	18122303	969
73	0.15	rs12085639	1	94478293	cg26904308	94477653	640
75	0.15	rs553046	7	140776401	cg11344121	140776653	252
79	0.15	rs986912	3	124284055	cg06218523	124284219	164
82	0.15	rs7166128	15	66946530	cg24382249	66947171	641
					cg09827761	66947392	862
85	0.15	rs17146982	7	22528568	cg10016647	22528033	535
88	0.15	rs10147026	14	62523148	cg25133706	62522891	257
					cg18261657	62523785	637
89	0.15	rs2244352	21	40757973	cg01322405	40757317	656
					cg00606841	40757691	282
					cg22858667	40757750	223
					cg26710963	40757899	74
					cg09916765	40758208	235
					cg21297395	40758325	352
					cg21459460	40758346	373
94	0.15	rs10499772	7	63766980	cg00834400	63767240	260
					cg09382800	63767753	773
					cg23849078	63767841	861
96	0.15	rs1870172	10	97990101	cg03281335	97989507	594
97	0.15	rs220030	15	25199768	cg21746532	25198793	975
					cg01432432	25199028	740
					cg08372135	25199057	711
					cg03858387	25199164	604
					cg02152271	25199270	498
					cg12298755	25199713	55
					cg18506672	25200253	485
					cg02125271	25200406	638
					cg26875073	25200490	722

99	0.14	rs10245302	7	70294323	cg18138580	70294264	59
					cg19560710	70294511	188
					cg25795398	70294525	202
100	0.14	rs2143346	6	26198449	cg23407396	26197481	968
					cg23601095	26197514	935
					cg16219491	26197733	716
					cg07329467	26199075	626

Supplementary Table 12

Table 4-19. Illumina 450K probes within 1kb of the top 100 cross-tissue ASM sites, as defined by average ASM score across individuals.

Rank	ASM	SNP	Chr	Mapinfo SNP	CpG site	Mapinfo CpG	Distance
1	0.24	rs10234308	7	22528590	cg10016647	22528033	557
2	0.22	rs12493005	3	176912899	cg03356026	176912588	311
3	0.21	rs1542180	2	177028767	cg13316854	177027856	911
					cg06145336	177027931	836
					cg25938806	177028162	605
					cg12634591	177028606	161
					cg01128482	177028621	146
					cg09387749	177028680	87
					cg23981871	177028714	53
					cg10304824	177028804	37
					cg00005847	177029073	306
					cg24541426	177029459	692
					cg05021643	177029608	841
8	0.18	rs4687210	3	191046770	cg25342125	191046380	390
					cg19066520	191046547	223
					cg23528751	191046721	49
					cg14972328	191046866	96
					cg00473167	191046872	102
					cg17596905	191046949	179
					cg23817170	191046997	227
					cg26325444	191047058	288
					cg09688588	191047460	690
11	0.17	rs1003533	5	131755651	cg03347934	131754928	723
14	0.17	rs220030	15	25199768	cg21746532	25198793	975
					cg01432432	25199028	740
					cg08372135	25199057	711
					cg03858387	25199164	604
					cg02152271	25199270	498
					cg12298755	25199713	55
					cg18506672	25200253	485
					cg02125271	25200406	638
					cg26875073	25200490	722
19	0.17	rs3121125	1	145382124	cg23685870	145382340	216
					cg07117663	145382963	839
21	0.16	rs17366322	2	11272784	cg15704155	11272399	385
					cg09045196	11272424	360
					cg10334385	11272868	84
					cg27641709	11272899	115
					cg05517905	11273075	291
					cg12453288	11273229	445
					cg03251967	11273291	507
25	0.16	rs3762199	20	18487792	cg26234223	18487771	21

					cg01866064	18487951	159
					cg27634549	18487976	184
					cg23490046	18487999	207
					cg15613982	18488017	225
					cg16156954	18488094	302
					cg07973146	18488185	393
					cg02081051	18488312	520
					cg22243844	18488366	574
					cg13313269	18488622	830
26	0.16	rs6713344	2	112814045	cg09071112	112813174	871
					cg06723287	112813613	432
					cg11931949	112814041	4
					cg19917083	112814452	407
29	0.16	rs3098382	5	71462632	cg20954533	71462729	97
30	0.16	rs17312325	X	68442113	cg00427681	68441389	724
					cg22337892	68441690	423
					cg03350825	68442327	214
37	0.15	rs1237543	13	24739364	cg03903831	24739249	115
38	0.15	rs7247675	19	28285519	cg17155612	28284698	821
					cg02942594	28284741	778
					cg03170611	28284813	706
					cg22473312	28284823	696
					cg08220598	28284952	567
					cg17871739	28284973	546
					cg24415565	28284988	531
					cg25246092	28285128	391
					cg03751272	28285308	211
					cg06949053	28285395	124
39	0.15	rs986912	3	124284055	cg06218523	124284219	164
40	0.15	rs7210715	17	7196390	cg00901493	7197375	985
42	0.15	rs7305595	12	3562234	cg23794848	3563020	786
46	0.15	rs2244352	21	40757973	cg01322405	40757317	656
					cg00606841	40757691	282
					cg22858667	40757750	223
					cg26710963	40757899	74
					cg09916765	40758208	235
					cg21297395	40758325	352
					cg21459460	40758346	373
47	0.15	rs7073504	10	60271987	cg07567376	60271505	482
					cg09696939	60272079	92
					cg07857251	60272754	767
					cg19940537	60272938	951
49	0.15	rs10883857	10	105127587	cg14125903	105126693	894
					cg10301212	105126702	885
					cg20188621	105127192	395
					cg11667101	105127581	6
					cg18586343	105127615	28
					cg20684973	105127632	45

					cg18732587	105127670	83
					cg25713684	105127687	100
					cg16500454	105127701	114
					cg01360325	105127811	224
					cg26978691	105128013	426
					cg00245789	105128183	596
					cg23418095	105128462	875
52	0.15	rs17146982	7	22528568	cg10016647	22528033	535
55	0.14	rs1139405	17	79478019	cg12793711	79477853	166
					cg26857837	79477974	45
					cg06713261	79478173	154
					cg15906692	79478715	696
					cg17317962	79478839	820
56	0.14	rs971649	8	124732275	cg13177458	124732088	187
60	0.14	rs2143346	6	26198449	cg23407396	26197481	968
					cg23601095	26197514	935
					cg16219491	26197733	716
					cg07329467	26199075	626
64	0.14	rs9331942	8	27455114	cg01572979	27455004	110
66	0.14	rs6874600	5	73945738	cg20658167	73944869	869
76	0.14	rs1673087	19	37264190	cg05616754	37263860	330
					cg03430717	37263925	265
					cg27080211	37263997	193
78	0.14	rs6548026	2	30872366	cg08120035	30872901	535
81	0.14	rs10499772	7	63766980	cg00834400	63767240	260
					cg09382800	63767753	773
					cg23849078	63767841	861
89	0.13	rs7185244	16	86546887	cg08142918	86546374	513
					cg03697918	86546628	259
					cg24908603	86546631	256
					cg07056644	86546785	102
					cg27453745	86546938	51
					cg09338251	86546979	92
					cg06834912	86547203	316
					cg04787888	86547322	435
					cg01243371	86547386	499
					cg00551679	86547530	643
					cg02783918	86547544	657
90	0.13	rs10454027	13	100259376	cg22853485	100258524	852
					cg12175106	100258589	787
					cg01726399	100258641	735
					cg10972431	100258753	623
					cg02168723	100258760	616
					cg27435467	100258954	422
					cg20863393	100259188	188
					cg24980005	100259219	157
					cg23862604	100259302	74
94	0.13	rs10108150	8	145687475	cg13071208	145686887	588

					cg17328964	145687451	24
					cg15189104	145687562	87
					cg02108130	145687727	252
					cg25390352	145688187	712
					cg22601688	145688206	731
98	0.13	rs8179387	1	22290156	cg27522078	22290005	151
99	0.13	rs9645876	13	112330861	cg06567836	112330930	69
					cg09984047	112330954	93
					cg15453778	112331089	228
					cg06996147	112331836	975

Supplementary Table 13

Table 4-20. Illumina 450K probes within 1kb of the top 100 tissue-specific ASM SNPs as defined by the high variability of ASM scores across tissues.

Rank	ASM SD	SNP	Chr	Mapinfo SNP	CpG site	Mapinfo CpG	Distance
2	0.15	rs927000	20	43598705	cg17104026	43598376	329
17	0.13	rs7959070	12	92805682	cg25400182	92806045	363
18	0.12	rs8032903	15	25980875	cg16651441	25980367	508
					cg02747151	25981177	302
					cg09978546	25981255	380
					cg14216870	25981322	447
23	0.12	rs6853895	4	38070899	cg24826020	38070998	99
24	0.12	rs7158663	14	101319424	cg19509303	101318703	721
					cg07460524	101319526	102
25	0.12	rs10783269	12	48995983	cg18467358	48995259	724
27	0.12	rs12510921	4	185189278	cg04960065	185189300	22
29	0.11	rs10481354	8	1704030	cg19131313	1704013	17
					cg17159058	1704075	45
31	0.11	rs1003533	5	131755651	cg03347934	131754928	723
33	0.11	rs2252397	7	140225750	cg01499591	140226367	617
38	0.11	rs12978286	19	17861868	cg26307105	17861017	851
					cg02639359	17862017	149
					cg08206267	17862104	236
					cg24783510	17862150	282
					cg11737334	17862210	342
					cg22317846	17862385	517
46	0.11	rs10411704	19	35800662	cg02776658	35800589	73
					cg22238209	35800743	81
					cg07918620	35800925	263
					cg15761414	35801014	352
49	0.11	rs1009014	6	158487143	cg18559896	158487541	398
					cg03544657	158487564	421
					cg07549551	158487608	465
51	0.10	rs3098382	5	71462632	cg20954533	71462729	97
56	0.10	rs9469529	6	33587443	cg08355863	33587496	53
					cg24603235	33588219	776
					cg19889152	33588278	835
					cg14639225	33588302	859
62	0.10	rs14067	13	114110660	cg05314639	114109729	931
					cg04870949	114109857	803
					cg03557916	114109949	711
64	0.10	rs11603592	11	20023522	cg09261702	20023953	431
70	0.10	rs7268068	20	62199258	cg23352030	62198469	789
					cg15262954	62198872	386
					cg17593958	62199034	224
					cg11779113	62199156	102
					cg06064964	62199181	77
					cg09844573	62199190	68
					cg17864091	62200091	833

					cg04300115	62200199	941
83	0.10	rs17539168	9	134485090	cg14485643	134484581	509
84	0.10	rs16843471	1	198590421	cg10768932	198590738	317
89	0.10	rs1774846	1	202994374	cg03639170	202993923	451
					cg00501542	202995246	872
95	0.10	rs10861495	12	106142981	cg23539745	106142108	873

Supplementary Table 14

Table 4-21. Illumina 450K probes within 1kb of the top 100 variable ASM sites, as defined by average range of ASM scores between individuals across tissues.

Rank	ASM Range	SNP	Chr	Mapinfo SNP	CpG site	Mapinfo CG	Distance
2	0.30	rs2244352	21	40757973	cg01322405	40757317	656
					cg00606841	40757691	282
					cg22858667	40757750	223
					cg26710963	40757899	74
					cg09916765	40758208	235
					cg21297395	40758325	352
					cg21459460	40758346	373
3	0.25	rs2346019	5	135415726	cg11852404	135414858	868
					cg00308130	135415190	536
					cg15837280	135415258	468
					cg07158503	135415693	33
					cg04515200	135415762	36
					cg13581155	135415781	55
					cg11978884	135415819	93
					cg11608150	135415948	222
					cg06478886	135416029	303
					cg04481923	135416205	479
					cg18678645	135416331	605
					cg06536614	135416381	655
					cg26328633	135416394	668
					cg25340688	135416398	672
					cg26896946	135416405	679
					cg00124993	135416412	686
					cg08745965	135416529	803
					cg16615357	135416594	868
					cg18797653	135416613	887
7	0.19	rs1209228	14	72231880	cg15926477	72231668	212
13	0.18	rs1695824	1	1365570	cg04865726	1365911	341
					cg12407057	1366206	636
					cg11156891	1366274	704
18	0.17	rs578597	9	104185258	cg14048797	104186172	914
23	0.16	rs3759917	15	92935447	cg18646851	92934809	638
					cg08152839	92936232	785
					cg05501584	92936306	859
26	0.16	rs11231646	11	63702635	cg05766107	63702235	400
28	0.16	rs1984713	4	47956556	cg06587443	47955755	801
					cg19000186	47956011	545
33	0.16	rs26835	16	2237549	cg07094340	2236773	776
					cg10118998	2236938	611
					cg07109743	2237020	529

					cg08938612	2237221	328
35	0.16	rs10952289	7	150524681	cg05405275	150525013	332
50	0.15	rs11611523	12	118018891	cg05738715	118018768	123
51	0.15	rs2400135	6	106253120	cg01182171	106252695	425
54	0.15	rs2377983	10	50352720	cg14620932	50353359	639
69	0.14	rs6779820	3	191046793	cg25342125	191046380	413
					cg19066520	191046547	246
					cg23528751	191046721	72
					cg14972328	191046866	73
					cg00473167	191046872	79
					cg17596905	191046949	156
					cg23817170	191046997	204
					cg26325444	191047058	265
					cg09688588	191047460	667
71	0.14	rs509062	20	56882238	cg21532801	56881949	289
76	0.14	rs8179356	1	100215981	cg17222434	100215737	244
80	0.14	rs7911580	10	7516306	cg03421195	7517215	909
84	0.14	rs2294490	1	15723843	cg00582948	15724194	351
87	0.14	rs4289794	8	141106204	cg12844366	141105335	869
					cg00085199	141106596	392
					cg05524387	141107113	909
91	0.14	rs6045166	20	18035837	cg02113429	18036013	176
94	0.13	rs12525580	6	74288782	cg05488681	74289464	682
98	0.13	rs3109398	3	182402101	cg09391284	182401187	914
					cg20830867	182402080	21

Supplementary Table 15

Table 4-22. Illumina 450K probes within 1kb of the top 100 variable ASM SNPs in whole blood as defined by range of ASM score between individuals.

Rank	Blood range	SNP	Chr	Mapinfo SNP	CpG site	Mapinfo CpG	Distance
3	0.25	rs509062	20	56882238	cg21532801	56881949	289
4	0.24	rs2244352	21	40757973	cg01322405	40757317	656
					cg00606841	40757691	282
					cg22858667	40757750	223
					cg26710963	40757899	74
					cg09916765	40758208	235
					cg21297395	40758325	352
					cg21459460	40758346	373
6	0.23	rs8179356	1	100215981	cg17222434	100215737	244
8	0.23	rs2346019	5	135415726	cg11852404	135414858	868
					cg00308130	135415190	536
					cg15837280	135415258	468
					cg07158503	135415693	33
					cg04515200	135415762	36
					cg13581155	135415781	55
					cg11978884	135415819	93
					cg11608150	135415948	222
					cg06478886	135416029	303
					cg04481923	135416205	479
					cg18678645	135416331	605
					cg06536614	135416381	655
					cg26328633	135416394	668
					cg25340688	135416398	672
					cg26896946	135416405	679
					cg00124993	135416412	686
					cg08745965	135416529	803
					cg16615357	135416594	868
					cg18797653	135416613	887
10	0.22	rs7694398	4	6733947	cg00966522	6734722	775
17	0.20	rs1890070	1	222628509	cg15078329	222628324	185
					cg11661340	222628632	123
					cg17836114	222628702	193
19	0.20	rs1984713	4	47956556	cg06587443	47955755	801
					cg19000186	47956011	545
23	0.19	rs10186949	2	237755950	cg04787916	237755762	188
30	0.19	rs849322	7	28331586	cg26353161	28331380	206
38	0.18	rs578597	9	104185258	cg14048797	104186172	914
46	0.18	rs165342	5	150884793	cg22672639	150884813	20
62	0.17	rs11231646	11	63702635	cg05766107	63702235	400
64	0.17	rs2072019	6	112460541	cg12188410	112460467	74
66	0.17	rs2357928	10	18549641	cg25091573	18549036	605
					cg24422029	18549378	263

					cg14716734	18549397	244
					cg05205813	18549413	228
					cg09987620	18549439	202
					cg26803268	18549536	105
					cg18959207	18549686	45
					cg02635932	18549778	137
					cg14357089	18550033	392
					cg26576398	18550135	494
					cg15720089	18550223	582
86	0.17	rs12938106	17	29926633	cg00351813	29926184	449
					cg25208821	29926356	277
89	0.17	rs7174808	15	52048251	cg02253142	52048211	40
91	0.17	rs1464172	3	124873295	cg17478992	124873768	473
					cg01432055	124873848	553
92	0.17	rs875981	11	130903093	cg01464703	130903135	42
96	0.17	rs10116248	9	132568500	cg14222752	132569372	872

Supplementary Table 16

Table 4-23. Illumina 450K probes within 1kb of the top 100 variable ASM SNPs in cerebellum as defined by range of ASM score between individuals.

Rank	Cerebellum range	SNP	Chr	Mapinfo SNP	CpG site	Mapinfo CpG	Distance
1	0.36	rs2244352	21	40757973	cg01322405	40757317	656
					cg00606841	40757691	282
					cg22858667	40757750	223
					cg26710963	40757899	74
					cg09916765	40758208	235
					cg21297395	40758325	352
					cg21459460	40758346	373
8	0.25	rs1695824	1	1365570	cg04865726	1365911	341
					cg12407057	1366206	636
					cg11156891	1366274	704
11	0.24	rs4289794	8	141106204	cg12844366	141105335	869
					cg00085199	141106596	392
					cg05524387	141107113	909
13	0.24	rs3759917	15	92935447	cg18646851	92934809	638
					cg08152839	92936232	785
					cg05501584	92936306	859
15	0.23	rs2346019	5	135415726	cg11852404	135414858	868
					cg00308130	135415190	536
					cg15837280	135415258	468
					cg07158503	135415693	33
					cg04515200	135415762	36
					cg13581155	135415781	55
					cg11978884	135415819	93
					cg11608150	135415948	222
					cg06478886	135416029	303
					cg04481923	135416205	479
					cg18678645	135416331	605
					cg06536614	135416381	655
					cg26328633	135416394	668
					cg25340688	135416398	672
					cg26896946	135416405	679
					cg00124993	135416412	686
					cg08745965	135416529	803
					cg16615357	135416594	868
					cg18797653	135416613	887
18	0.23	rs941483	14	94422623	cg00839873	94421989	634
					cg18572214	94423192	569
					cg01956154	94423399	776
28	0.21	rs578597	9	104185258	cg14048797	104186172	914
35	0.20	rs2377983	10	50352720	cg14620932	50353359	639
40	0.19	rs1209228	14	72231880	cg15926477	72231668	212

42	0.19	rs7426865	3	149058065	cg06305422	149057820	245
60	0.18	rs12412942	10	2357702	cg23939236	2357224	478
					cg03163525	2357283	419
					cg17246535	2357363	339
					cg03946168	2357461	241
					cg18730194	2357587	115
68	0.18	rs1708337	3	124886822	cg24838136	124887734	912
76	0.17	rs4253747	22	46613237	cg06467025	46614169	932
77	0.17	rs17146982	7	22528568	cg10016647	22528033	535
88	0.17	rs9348131	6	166850798	cg11314174	166850928	130
					cg02624984	166851100	302
					cg21313980	166851372	574
90	0.17	rs26835	16	2237549	cg07094340	2236773	776
					cg10118998	2236938	611
					cg07109743	2237020	529
					cg08938612	2237221	328
96	0.17	rs7911580	10	7516306	cg03421195	7517215	909
97	0.17	rs3735715	8	102679909	cg25876227	102680096	187

Supplementary Table 17

Table 4-24. Illumina 450K probes within 1kb of the top 100 variable ASM SNPs in cortex (BA9) as defined by range of ASM score between individuals.

Rank	BA9 range	SNP	Chr	Mapinfo SNP	CpG site	Mapinfo CpG	Distance
3	0.30	rs2346019	5	135415726	cg11852404	135414858	868
					cg00308130	135415190	536
					cg15837280	135415258	468
					cg07158503	135415693	33
					cg04515200	135415762	36
					cg13581155	135415781	55
					cg11978884	135415819	93
					cg11608150	135415948	222
					cg06478886	135416029	303
					cg04481923	135416205	479
					cg18678645	135416331	605
					cg06536614	135416381	655
					cg26328633	135416394	668
					cg25340688	135416398	672
					cg26896946	135416405	679
					cg00124993	135416412	686
					cg08745965	135416529	803
					cg16615357	135416594	868
					cg18797653	135416613	887
4	0.29	rs2244352	21	40757973	cg01322405	40757317	656
					cg00606841	40757691	282
					cg22858667	40757750	223
					cg26710963	40757899	74
					cg09916765	40758208	235
					cg21297395	40758325	352
					cg21459460	40758346	373
9	0.24	rs1209228	14	72231880	cg15926477	72231668	212
10	0.23	rs9357087	6	11216700	cg05208582	11216594	106
11	0.22	rs503755	2	135228060	cg05670240	135227167	893
22	0.20	rs17008180	4	106406083	cg08987594	106405819	264
					cg20364776	106405850	233
33	0.19	rs2377983	10	50352720	cg14620932	50353359	639
35	0.19	rs9395212	6	46726441	cg10669279	46726850	409
47	0.18	rs633644	21	44783063	cg02260098	44782434	629
					cg25191041	44782470	593
					cg14081667	44782497	566
					cg19318330	44782872	191
49	0.18	rs985462	10	127769598	cg19997196	127769491	107
					cg08295661	127769903	305
51	0.18	rs10952289	7	150524681	cg05405275	150525013	332
54	0.18	rs17573298	16	78218016	cg04344190	78218445	429
56	0.18	rs6577101	13	113293717	cg14227911	113292824	893

					cg07014973	113292951	766
57	0.18	rs12525580	6	74288782	cg05488681	74289464	682
58	0.18	rs4784910	16	58298471	cg08638917	58299153	682
72	0.18	rs17130554	1	68653137	cg22850128	68654112	975
76	0.18	rs17432433	15	37174748	cg17921924	37174376	372
					cg22444584	37174661	87
					cg27042266	37175007	259
					cg16890051	37175029	281
					cg12288941	37175117	369
90	0.17	rs10886631	10	85325650	cg12554194	85324700	950
					cg04567661	85325001	649
					cg24127335	85325062	588
92	0.17	rs11135514	5	96790262	cg21913787	96789385	877
98	0.17	rs2070682	7	100777267	cg17968347	100777740	473

Supplementary Table 18

Table 4-25. Sample information. Shown is demographic information for individuals profiled for ASM using our MSNP protocol and for DNA methylation using the Illumina Infinium HumanMethylation450 BeadChip (450K). PMD = Post-mortem delay (h).

	Demographics			Tissue							
ID	Sex	Age at death	PMD	BA8	BA9	BA10	BA 17	BA21	BA28/34	Cerebellum	Blood
1	F	82	43	MSNP	MSNP	MSNP		MSNP	MSNP	MSNP	MSNP
					450K			450K	450K	450K	450K
2	M	78	NA	MSNP	MSNP	MSNP			MSNP	MSNP	MSNP
					450K			450K	450K	450K	450K
3	F	92	17	MSNP	MSNP	MSNP	MSNP	MSNP	MSNP	MSNP	MSNP
					450K			450K	450K	450K	450K
4	F	82	13		450K			450K	450K	450K	450K
5	F	81	17		450K			450K	450K	450K	
6	M	86	7		450K			450K	450K	450K	450K
7	F	80	3		450K			450K	450K	450K	
8	F	55	12		450K			450K	450K	450K	
9	M	79	47		450K			450K	450K	450K	
10	M	80	21		450K			450K	450K	450K	
11	M	86	6		450K			450K	450K	450K	
12	F	87	22		450K			450K		450K	
13	F	68	9					450K		450K	
14	M	59	50		450K			450K	450K	450K	
15	F	80	48		450K			450K	450K	450K	
16	M	40	40					450K		450K	
17	F	55	95		450K			450K		450K	
18	F	73	70		450K			450K			
19	M	80	60		450K			450K	450K	450K	
20	F	90	50		450K			450K	450K	450K	

21	M	81	18	450K	450K	450K	450K	
22	M	66	52	450K	450K	450K	450K	
23	F	88	6	450K	450K	450K	450K	
24	M	86	21	450K	450K	450K	450K	450K
25	F	90	56	450K	450K	450K	450K	450K
26	F	92	23	450K	450K	450K	450K	450K
27	M	97	44	450K	450K	450K	450K	
28	M	74	23	450K	450K	450K	450K	
29	M	77	11	450K	450K	450K	450K	
30	M	82	47	450K	450K	450K	450K	
31	M	73	23	450K	450K	450K	450K	
32	M	95	26	450K	450K	450K	450K	
33	F	79	56	450K	450K	450K	450K	
34	F	91	74	450K	450K	450K	450K	
35	M	NA	NA					450K
36	F	NA	NA					450K
37	F	NA	NA					450K
38	F	NA	NA					450K
39	M	NA	NA					450K

Supplementary Table 19

Table 4-26. Number of informative loci interrogated by the MSNP protocol for each sample.

	BA8	BA9	BA10	BA17	BA21	BA28/34	Cerebellum	Blood	Total
Individual 1	110004	110004	110004		110004	110004	110004	110004	110004
Individual 2	107649	107649	107649			107649	107649	107649	107649
Individual 3	108942	108942	108942	108942	108942	108942	108942	108942	108942
Total	220449	220449	220449	108942	220449	177162	220449	220449	

Supplementary Table 20

Table 4-27. Primers used for clonal bisulfite sequencing.

Assay	Forward Primer	Reverse Primer
SNP_A-4255628 (rs959246)	TGGTTTGGTAGGTGTTAGGTATTTAG	CTCCCTATTCCCCAATAAAAATTTA
SNP_A-2273834 (rs2252267)	TTTTTGAGTGATTGGTTATTAGAAA	CCCAAAAAAACTAAAATTCCCTAAC

5. A genome-wide study of H3K27ac identifies molecular signatures of Alzheimer's disease pathology

5.1 Introduction

Alzheimer's disease (AD) is a chronic neurodegenerative disorder that contributes substantially to the global burden of disease, affecting in excess of 26 million people worldwide (Brookmeyer et al., 2007, Knapp et al., 2007). AD is characterized by cognitive decline and memory loss, and associated with progressive neuropathology in the cortex, with areas surrounding the entorhinal cortex being particularly affected early in the disease (Wenk, 2003). The neuropathological hallmarks of the disease include the extracellular deposition of neurotoxic amyloid-β peptides in the form of amyloid plaques and an accumulation of intracellular neurofibrillary tangles composed of hyperphosphorylated tau (Hardy and Selkoe, 2002, Hardy and Higgins, 1992). Despite progress in understanding risk factors contributing to AD progression, the mechanisms involved in disease progression are not fully understood and long-term treatments, reversing the cellular disease process, are elusive.

There has been considerable success in identifying genetic risk factors for AD (Karch et al., 2014). While autosomal dominant mutations in three genes (*APP*, *PSEN1*, and *PSEN2*) can explain early-onset (< 65 years) familial AD, these account for only 1-5% of the total disease burden (Reitz et al., 2011). Most cases of AD are late-onset (> 65 years), non-Mendelian and highly sporadic, with susceptibility attributed to the action of highly prevalent genetic variants of low penetrance. In addition to the well-established risk associated with the *APOE* locus (Corder et al., 1993, Bu, 2009, Liu et al., 2013) there has been notable success in identifying novel AD-associated variants capitalising on the power of genome-wide association studies (GWAS) in large sample cohorts; a recent large GWAS meta-analysis of AD, incorporating 74,046 samples, identified 19 genome-wide significant risk loci for sporadic AD (Lambert et al., 2013). Despite these advances, little is known about the functional mechanisms by which risk variants mediate disease susceptibility.

Increased understanding about the functional complexity of the genome has led to growing recognition about the likely role of non-sequence-based regulatory variation in health and disease. Building on the hypothesis that epigenomic dysregulation is important in the aetiology and progression of neuropathology (Lunnon and Mill, 2013), we and others have recently performed the first genome-

scale cross-tissue analyses of DNA methylation in AD identifying robust DNA methylation differences associated with AD neuropathology across multiple independent human post-mortem brain cohorts (Lunnon et al., 2014, De Jager et al., 2014). To date, however, no study has systematically examined other types of regulatory genomic variation in AD. In this study, we focus on the histone modification H3K27ac, a robust mark of active enhancers and promoters that is strongly correlated with gene expression and transcription factor binding (Creyghton et al., 2010, Rada-Iglesias et al., 2011, Spitz and Furlong, 2012). In addition, H3K27ac is characterized by elevated levels of inter-individual variability compared to other modifications, in particular when located in enhancer regions (Kasowski et al., 2013), making it an appropriate target for epidemiological profiling. Interestingly, histone deacetylase (HDAC) inhibitors have been shown to ameliorate symptoms of cognitive decline and synaptic dysfunction in mouse models of AD (Kilgore et al., 2010, Graff and Tsai, 2013b, Cuadrado-Tejedor et al., 2017) and are promising targets for novel human AD treatments (Graff and Tsai, 2013a, Fischer, 2014, Lu et al., 2015). Despite this, investigations into alterations of the global levels of histone acetylation in AD have thus far been inconclusive (Rao et al., 2012, Zhang et al., 2012, Lu et al., 2014, Narayan et al., 2015). Few studies have systematically profiled H3K27ac across large numbers of samples in the context of complex disease, and optimal methods for these analyses are still being developed (Sun et al., 2016).

We used chromatin immunoprecipitation combined with highly-parallel sequencing (ChIP-seq) to profile H3K27ac in post-mortem entorhinal cortex samples from AD patients and matched controls. We identify robust regulatory genomic signatures associated with AD, including variable H3K27ac across discrete regions annotated to genomic loci mechanistically implicated in the onset of both tau and amyloid pathology. This is the first study of variable H3K27ac yet undertaken for AD, and in addition to identifying molecular pathways associated with AD neuropathology, we present a framework for genome-wide studies of this modification in complex disease

5.2 Methods

5.2.1 Samples

Post-mortem brain samples from 54 individuals - 27 characterized by Alzheimer's disease neuropathology and 27 neuropathology-free brain samples - were provided by the MRC London Neurodegenerative Disease Brain Bank (http://www.kcl.ac.uk/ioppn/depts/cn/research/MRC-London-Neurodegenerative-Diseases-Brain-Bank/MRC-London-Neurodegenerative-Diseases-Brain-Bank.aspx). Ethical approval for the study was provided by the NHS South East London REC 3. Samples for this ChIP-seq study were selected from a larger collection of post-mortem entorhinal cortex (Brodmann area (BA) 28/34) samples, based on Braak staging, a standardized measure of neurofibrillary tangle burden determined at autopsy (Braak and Braak, 1991). We prioritized cases with high Braak staging and controls with lower Braak scores (**Table 5-1**). Subjects were approached in life for written consent for brain banking, and all tissue donations were collected and stored following legal and ethical guidelines (NHS reference number 08/MRE09/38; the HTA license number for the LBBND brain bank is 12293). All samples were dissected by trained specialists, snap-frozen and stored at −80 °C. A detailed list of demographic and sample data for each individual included in the final analyses is provided in **Table 5-1**.

Table 5-1. Demographic and phenotypic information for each individual sample included in differential acetylation profiling.

Sample	Post-mortem diagnosis	Braak stage	Age at death	Sex	Neuronal proportion estimate*	Post-mortem delay (mins)
1	Control	0	63	M	12%	1380
2	Control	0	84	M	36%	3180
3	Control	0	69	F	33%	2880
4	Control	0	77	F	8%	1260
5	Control	0	66	F	36%	3000
6	Control	0	90	M	20%	2700
7	Control	0	66	F	48%	4680
8	Control	0	58	M	30%	3480
9	Control	1	72	F	22%	2820

10	Control	1	65	M	18%	1560
11	Control	1	67	M	46%	1500
12	Control	2	80	M	45%	3300
13	Control	2	76	F	41%	1320
14	Control	2	85	F	37%	2700
15	Control	2	67	F	15%	3720
16	Control	2	80	M	45%	1860
17	Control	2	68	M	29%	3600
18	Control	2	85	M	33%	3300
19	Control	2	83	F	NA	2340
20	Control	2	89	F	27%	2580
21	Control	3	97	M	20%	2460
22	Control	3	88	F	28%	2340
23	Control	3	78	M	36%	1440
24	AD	6	91	M	1%	4080
25	AD	6	75	F	34%	720
26	AD	6	84	M	32%	4020
27	AD	6	65	F	10%	4200
28	AD	6	62	M	40%	2940
29	AD	6	74	M	21%	930
30	AD	6	73	F	18%	5460
31	AD	6	86	F	17%	1260
32	AD	6	70	M	13%	5400
33	AD	6	80	M	37%	2700
34	AD	6	87	F	26%	1860
35	AD	6	68	M	12%	1620
36	AD	6	71	M	23%	1260
37	AD	6	86	F	16%	660
38	AD	6	94	F	30%	2520
39	AD	6	83	M	21%	5760
40	AD	6	76	M	30%	3300
41	AD	6	87	F	17%	2400
42	AD	6	62	M	17%	1380
43	AD	6	83	M	42%	4260
44	AD	6	79	F	8%	3780
45	AD	6	81	F	11%	1200
46	AD	6	88	F	33%	1080
47	AD	6	81	F	60%	4800

* Neuronal proportion estimates are derived from DNA methylation data generated on the same samples using the Illumina 450K HumanMethylation Array, calculated using the *CETS* R package (Guintivano et al., 2013).

5.2.2 Chromatin immunoprecipitation

30 mg brain samples were homogenized in 1ml ice-cold phosphate buffered saline (PBS) buffer with protease inhibitor cocktail (PIC; Diagenode, Cat #C12010012). The suspension was centrifuged at 4,000 rpm for 5 minutes at 4°C, discarding the supernatant. The pellets were resuspended in 1ml PBS containing 1% formaldehyde, rotating at room temperature for 8 minutes. The cross-linking process was terminated by adding 100µl glycine, followed by 5 minutes of rotation. After 5 minutes of centrifugation at 4,000 rpm and 4°C, the pellet was washed twice with ice-cold PBS (suspend the pellet in 1ml PBS with PIC, centrifuge for 5 minutes at 4,000 rpm and 4°C, and discard the supernatant), then lysed in 10 mL ice-cold lysis buffer iL1 and iL2 (iDeal ChIP-seq kit; Diagenode, Cat #C01010050, Denville, NJ, USA), sequentially (suspend the pellet in 10ml lysis buffer, incubate mixing gently for 10 minutes at 4°C, centrifuge for 5 minutes at 4,000 rpm and 4°C, and discard the supernatant). The cross-linked lysate was suspended in 1.8ml shearing buffer iS1 (iDeal ChIP-seq kit; Diagenode, Cat #C01010050, Denville, NJ, USA) containing PIC and sonicated in aliquots of 300µl for 10 minutes on a Bioruptor Pico (Diagenode, Denville, NJ, USA) with 30 seconds on/off cycles. After shearing, samples were centrifuged at 14,000 rpm for 10 minutes, collecting the supernatant, containing the sheared chromatin. Sonication resulted in an average size range of 100-1000bp as visualized by agarose gel electrophoresis (**Figure 5-1**).

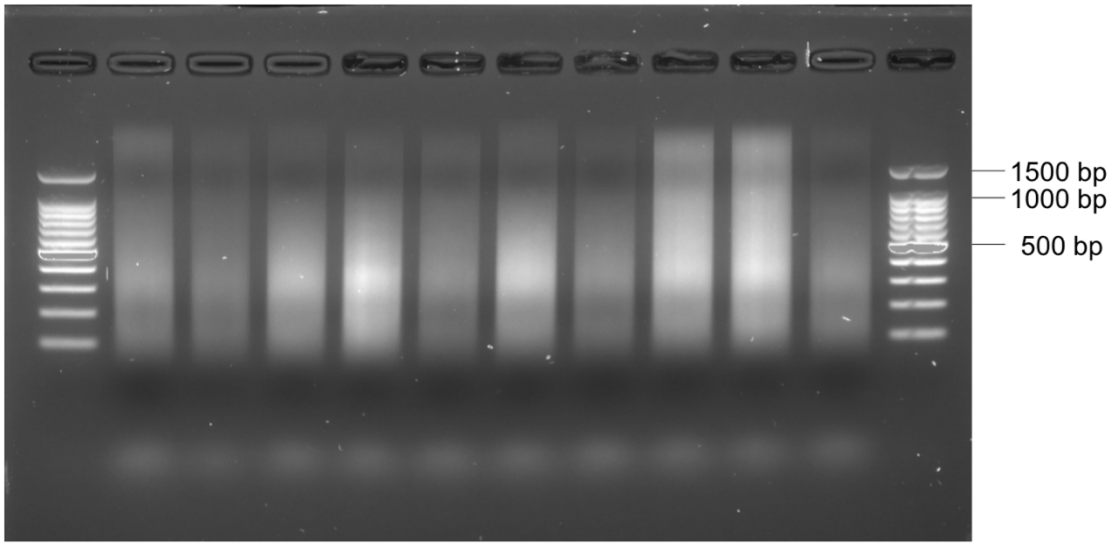


Figure 5-1. Chromatin shearing resulted in an average size distribution of 100-1000bp. 10 example agarose gel electrophoresis profiles, generated from DNA extracted after sonication, are shown.

Immunoprecipitation was performed using the iDeal ChIP-seq kit (Diagenode, Cat #C01010050, Denville, NJ, USA) on the SX-8G IP-Star robot (Diagenode, Liège, Belgium), following the manufacturer's protocol. All samples were immunoprecipitated with H3K27ac polyclonal antibody (Diagenode, Cat #C15410196, Liège, Belgium). In addition, a randomly selected subgroup of 12 samples – 6 cases and 6 controls – were immunoprecipitated with rabbit IgG antibody (iDeal ChIP-seq kit; Diagenode, Cat #C01010050, Denville, NJ, USA) as negative control. 1-1.5µl of H3K27ac or IgG antibody were first mixed with 98.5-99µl ChIP buffer iC1 (iDeal ChIP-seq kit; Diagenode, Cat #C01010050, Denville, NJ, USA), 0.5µl PIC and 4µl of 5% bovine serum albumin (BSA; iDeal ChIP-seq kit; Diagenode, Cat #C01010050, Denville, NJ, USA), which was incubated with magnetic beads for 3 hours at 4°C. Next the antibody conjugate was added to 180µl chromatin for overnight (15h) immunoprecipitation at 4°C in an immunoprecipitation mix also containing 20µl ChIP buffer iC1, 1µl PIC and 4µl of 5% BSA. After immunoprecipitation, the beads were suspended in in 100µl elution buffer iE1 (iDeal ChIP-seq kit; Diagenode, Cat #C01010050, Denville, NJ, USA), to which 4µl elution buffer iE2 (iDeal ChIP-seq kit; Diagenode, Cat #C01010050, Denville, NJ, USA) was added. Cross-link reversal was performed on a PCR thermoblock for 4 hours at 65°C. DNA was extracted using columns (Diagenode Micro ChIP DiaPure columns, Cat No C03040001) according to the manufacturers protocol, eluding the DNA from the column matrix in 30µl DNA elution buffer (MicroChIP DiaPure columns; Diagenode, Cat #C03040001, Denville, NJ, USA). Quantitative PCR, using 1% input DNA, confirmed specific enrichment of H3K27ac at positive control genes (*IEF4A2* and *GAPDH*) but not at negative control genes (*MBex2* and *TSH2b*; all primers were provided by Diagenode).

5.2.3 Sequencing

Libraries were prepared using the MicroPlex Library Preparation kit (Diagenode, Cat #C05010010, Denville, NJ, USA) according to the manufacturer's protocol. DNA concentrations were measured on Qubit dsDNA HS Assay Kits (Invitrogen, Cat# Q32851, Carlsbad, CA, USA) on the Qubit 2.0 Fluorometer (Invitrogen) and library fragment profiles generated on the Agilent 2100 BioAnalyzer using Agilent

High Sensitivity DNA kits (Agilent Technologies, Cat# 5067-4626, Santa Clara, CA, USA). 7 samples were excluded from sequencing based on poor qPCR results after immunoprecipitation or low library concentration and a sample mismatch. The remaining 47 samples (from 24 cases and 23 controls) were sequenced on an Illumina HiSeq-2500 using single-end sequencing and a read length of 50bp.

5.2.4 Data pre-processing and quality control

Global sample anomalies were ruled out using *fastqc* (Andrews, 2010) summary measures. All fastq files were aligned to the *homo sapiens* reference genome (hg19, Broad Institute) using *Bowtie* (Langmead et al., 2009). The output SAM files were converted to binary (BAM) format. All BAM files were sorted and indexed using *samtools* (Li et al., 2009). PCR duplicates were removed using *Picard* (http://broadinstitute.github.io/picard/). *Samtools* was used to additionally remove non-uniquely mapped reads as well as reads with a sequencing quality score q < 30. Final read counts after QC for all 47 samples are shown in **Figure 5-2.** On average, we obtained 30,032,623 reads per sample (SD= 10,638,091; range=10,910,000-53,770,000) and individual read counts did not associate with disease status (P = 0.93).

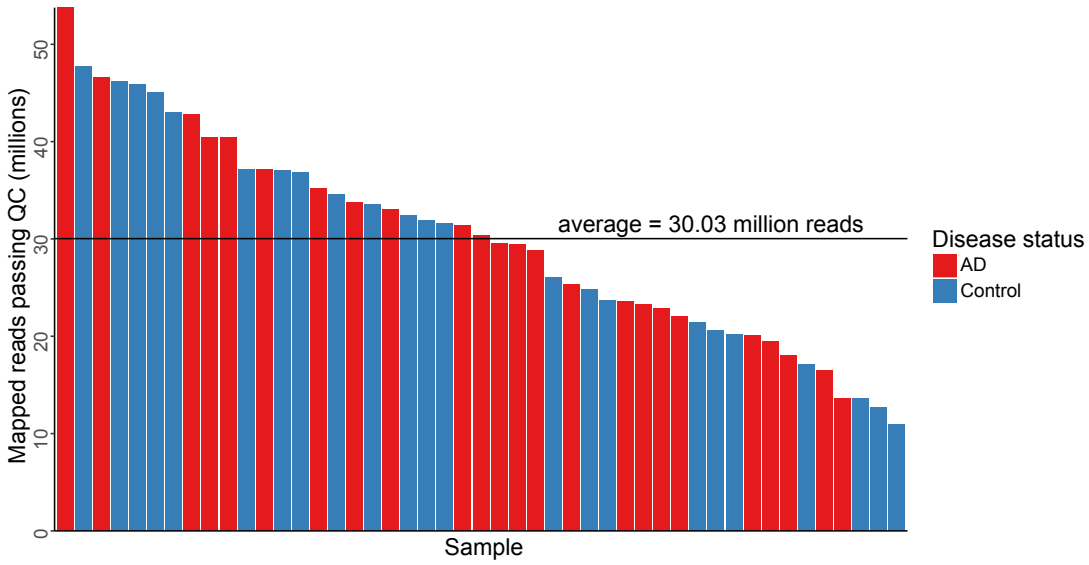


Figure 5-2. The number of ChIP-seq reads passing QC for each sample. On average, we obtained 30,032,623 reads per sample (SD = 10,638,091; range = 10,910,000-53,770,000) after

filtering and QC. There was no significant difference in the number of sequencing reads obtained for AD cases (red) and controls (blue) (P = 0.93).

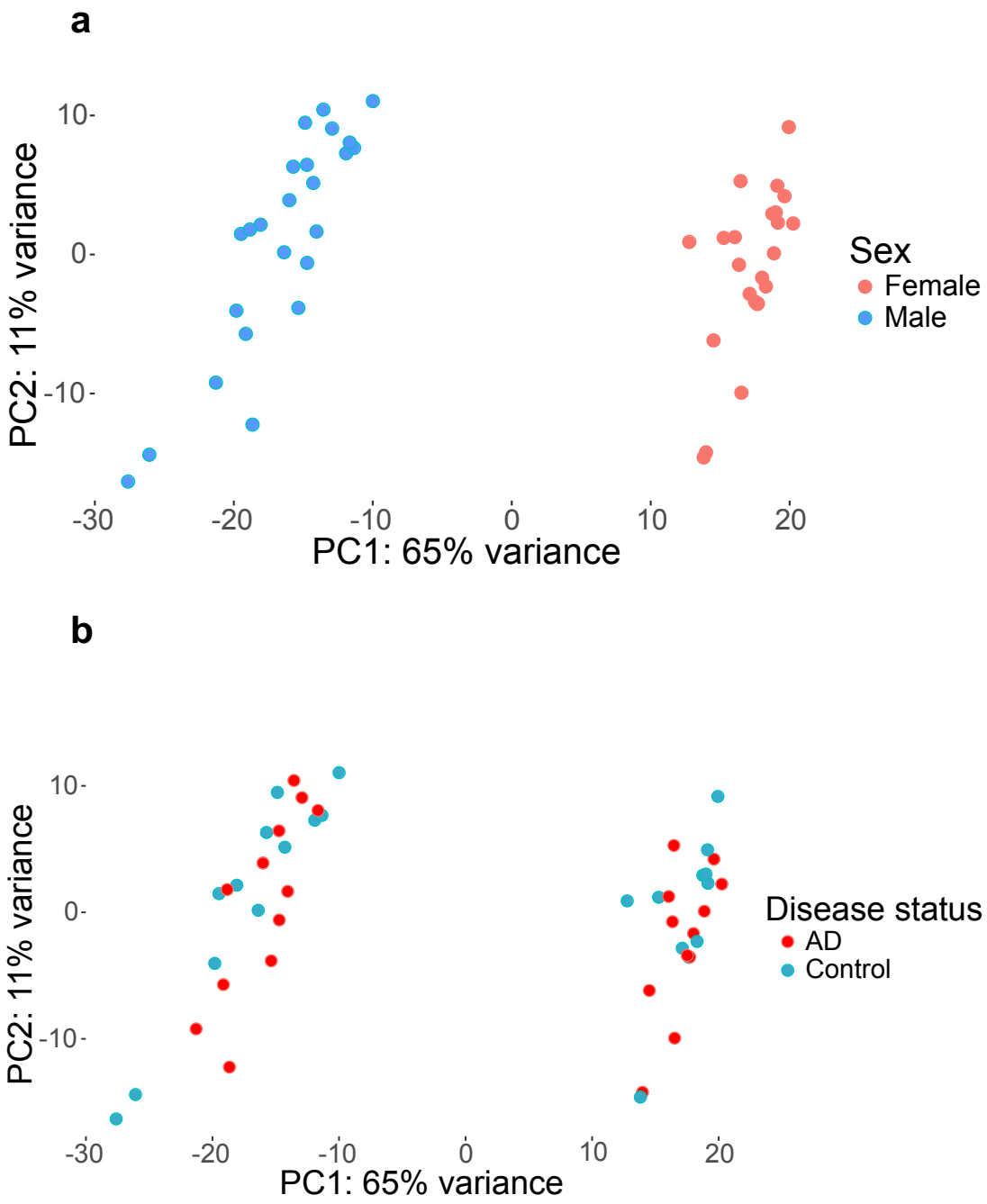


Figure 5-3. PCA analysis confirms documented sex. (**a**) Load on the first principal components calculated on normalised read counts shows exact overlap in annotated sex and epigenetically predicted sex (1st PC). (**b**) Variance captured by the same first PCs is not associated with disease status.

5.2.5 Peak calling and read counts

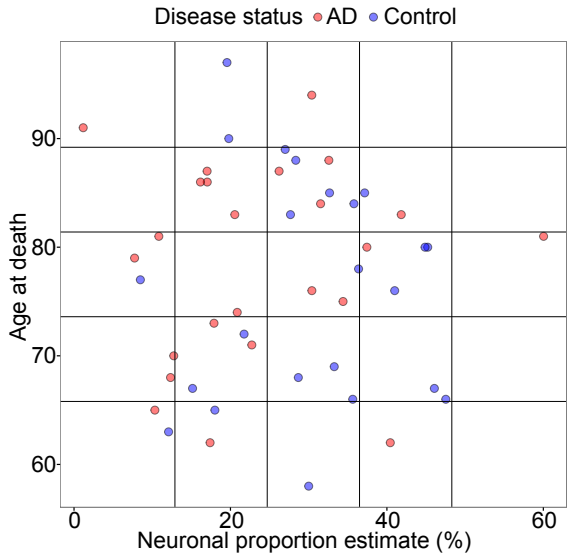
All filtered BAM files were merged into one grouped file and converted to tagAlign format using *bedtools* (Quinlan and Hall, 2010). Peaks were called on this

merged file using *MACS2* (Zhang et al., 2008), keeping all duplicates, since duplicates were removed from each sample previously and any remaining duplicates would result from the same read occurring in more than one sample. From the resulting peaks those located in unmapped contigs and mitochondrial DNA were filtered out as well as peaks that did not meet a significance threshold of $P < 1.0 \times 10^{-7}$ for peak calling. The bed file of peaks was converted to gff format using awk and R and reads for each individual sample were generated using *HTSeq* (Anders et al., 2015). Final filtering was performed using the Bioconductor package *EdgeR (*Robinson et al., 2010*)*, excluding peaks with fewer than 2 samples showing at least 1 read per million, resulting in a total of 182,065 peaks to be tested. Principal components analysis (PCA) in R using *DESeq2* (Love et al., 2014) confirmed that the epigenetically predicted gender was identical to the recorded one (**Figure 5-3**) and load on the first two principal components was not related to the sample's disease status.

5.2.6 Peak validation

We validated the 182,065 union peaks in two ways. First, we obtained the locations of H3K27ac peaks called in the cortex (BA9) and cerebellum from a recent paper by Sun and colleagues (Sun et al., 2016). Second, we downloaded H3K27ac profiles produced by the NIH Roadmap Epigenomics Consortium (Kundaje et al., 2015) from the Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo) for a range of cell-/tissue-types including several brain regions (mid frontal lobe (GSM773015), inferior temporal gyrus (GSM772995), middle hippocampus (GSM773020), substantia nigra (GSM997258), cingulate gyrus (GSM773011), H1-derived neuronal progenitor cells (HDNPs, GSM753429), lung (GSM906395), liver (GSM1112808) and skeletal muscle (GSM916064)). The downloaded files were in bed format, on which we performed peak calling using MACS2 and the same specifications as described for our own samples, discounting any duplicate reads. We calculated the overlap between each peak set and our peaks by quantifying the percentage of peaks from the external sample overlapping our peaks using the Bioconductor package *GenomicRanges* (Lawrence et al., 2013).

a



b

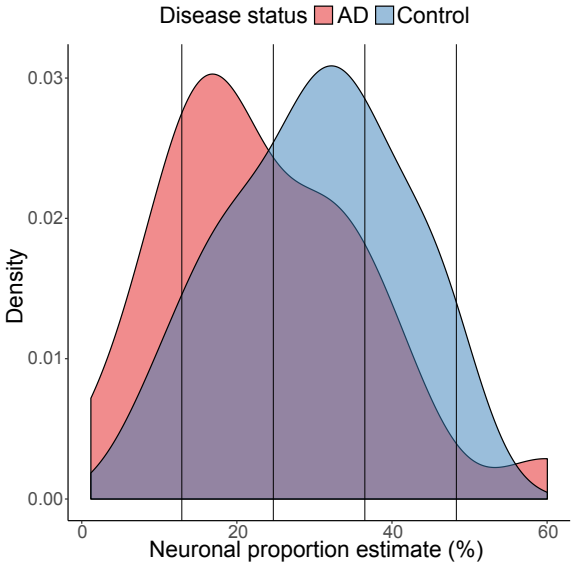


Figure 5-4. Factor-conversion of age and neuronal proportion estimates. Neuronal cell proportion estimates (derived from DNA methylation data using *CETS* (Guintivano et al., 2013)) and age at death were factorised, to adapt for use in the *EdgeR* statistical model. (**a**) The distribution of the samples' age at death and neuronal proportion estimates based on CETS profiles are plotted together with the four breaks generated for the respective factor versions of the variable by the cut() function in R. (**b**) The distribution of the CETS variable highlights decreased neuronal proportion estimates in cases (AD) (average = 23.8%) compared to controls (average = 30.2%).

5.2.7 Differential Peak calling

We used the quasi-likelihood F test (Lun et al., 2016) in *EdgeR* (Robinson et al., 2010) to analyse peak differences between AD-cases and controls, allowing us

to correct for potential confounders in the analysis of differential peaks. Our analyses accounted for additional phenotypic variation across the samples, including age at death and neuronal proportion estimates (CETS) based on DNA methylation profiles from the Illumina 450K HumanMethylation Array from the same samples, which were calculated using the *CETS* R package (Guintivano et al., 2013). We imputed the median CETS estimate for one individual with missing DNA methylation data. To include age at death and CETS estimate in the *EdgeR* differential peak calling method, these variables were converted to five-level factors using the R function *cut()*. The distribution of the age and CETS variable (including the imputed individual) with the respective bins of the factor variables are shown in **Figure 5-4**. We next calculated normalization factors based on sample-specific library compositions and estimated both sample and peak-specific dispersions, specifically for a generalized linear model controlling for factorized CETS estimates and age at death. The quasi-likelihood F-test was conducted after fitting a quasi-likelihood model (Lun et al., 2016) using the *glmQLFit()* and *glmQLFTest()* functions respectively. We generated genome browser tracks on the UCSC Genome Browser using BAM files from three selected AD samples and three controls, as well as bed tracks for background and significantly hyper- and hypoacetylated peaks (FDR < 0.05). Effect sizes are reported as log fold change, a standard measure for quantifying sequencing read count differences between different conditions. Log fold change refers to the $\log_2$-transformed ratio of normalized read counts between cases and controls, with positive values indicating higher normalized read counts in AD samples.

5.2.8 Analysis of CpG island peaks and genomic locations

First, we analysed the sequence content of all H3K27ac peaks using the hg19 reference genome sequence. For each peak we quantified the GC content, as well as the observed vs expected number of CpG sites, defining peaks with > 50% GC content and > 60% observed over expected CpG sites as CpG island peaks (Gardiner-Garden and Frommer, 1987). Second, we downloaded the location of all transcription start sites (TSS) of RefSeq genes in the hg19 assembly and subsequently compared peaks that mapped within 2kb upstream of a TSS to those that didn't. For both classification methods we compared the

distributions of differentially acetylated peaks to those of the non-significant background peaks using Fisher's exact tests.

5.2.9 Analysis of copy-number variation (CNV) and short tandem repeats (STRs) across H3K27ac peaks

We downloaded published maps of CNVs (Zarrei et al., 2015) and STRs (Willems et al., 2014) to analyse the enrichment of these amongst differentially acetylated peaks. We overlapped the location of all H3K27ac peaks with these locations and tested for statistical enrichments of CNVs and STRs using Fisher's exact test.

5.2.10 Genomic annotation and enrichment analyses

Peaks were annotated to genes using the *Genomic Region Enrichment and Annotation Tools* (*GREAT*) (McLean et al., 2010). In addition, we performed enrichment analyses to calculate statistical enrichments for ontological annotation (gene ontologies for molecular function and biological processes (Ashburner et al., 2000), human diseases (Osborne et al., 2009) as well as human phenotypes (Robinson and Mundlos, 2010)). Functional enrichment analyses were conducted for significantly hyper- and hypoacetylated peaks (FDR < 0.05) separately, using the basal plus extension option. Significance in the enrichment test is based on a hypergeometric test of genes annotated to the test set (hyper-/ hypoacetylated peaks) compared to the background set of genes annotated to all 182,065 peaks called across all samples. Results presented in **Figure 5-19** are restricted to the top five non-redundant enrichments (separated by at least two nodes in the local directed acyclic graph visualizing the hierarchy of enriched terms from a single ontology) associated with at least three genes in the test set for the ontology categories biological process, molecular function, and disease ontology and we show full enrichments across all categories in **Appendix C**.

5.2.11 Motif enrichment analysis

Motif analysis was performed using the *Regulatory Sequence Analysis Tools suite* (*RSAT*) (Thomas-Chollier et al., 2012b, Thomas-Chollier et al., 2012a) at http://rsat.sb-roscoff.fr. Peak sequences were reduced to 500bp to each side of the peak centre, and motif discovery was conducted on 6 and 7mer oligonucleotides, comparing the statistically enriched sequences with known transcription factor motifs from JASPAR (Mathelier et al., 2016) (core nonredundant vertebrates) and Homer (Heinz et al., 2010) (Human TF motifs). Enrichments were computed relative to the background peak sequences (n = 182,065 peaks) for significantly hyper- and hypoacetylated peaks (FDR < 0.05).

5.2.11 Analysis of differential H3K27ac across AD regions from genome-wide association studies (GWAS):

The summary statistics for the stage 1 GWAS from (Lambert et al., 2013) were downloaded from http://web.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php. These results were clumped (p1 = 1e-4; p2 = 1e-4, r2 =0.1, window = 3000kb) using plink, which collapses multiple correlated signals (due to linkage disequilibrium (LD)) into regions which represent independent signals. LD relationships were inferred from a reference GWAS dataset from another study (Phase 1 from (Hannon et al., 2016a)). Neighbouring regions located within 250kb of each other on the same chromosome were subsequently merged. After clumping, each region was assigned the minimum *P* value for all SNPs contained in the region (from Lambert et al), and regions were then filtered to the genome-wide significance threshold (*P* < 5.0E-08). This yielded 11 LD blocks for the genome-wide significant findings from Lambert et al., which were then overlapped with our AD-associated differentially acetylated peaks using the Bioconductor package *GenomicRanges* (Lawrence et al., 2013).

5.2.12 Integrative analysis with DNA methylation and hydroxymethylation

DNA methylation and hydroxymethylation data was available (unpublished) from entorhinal cortex DNA for 42 of the samples profiled in this study. DNA methylation and hydroxymethylation profiles were generated on the Illumina Infinium HumanMethylation450 BeadChip (Illumina Inc., CA, USA) ("Illumina 450K array") using the TrueMethyl Array kit (Cambridge Epigenetix, Cambridge, UK). Profiles for both modifications were pre-processed, normalized and filtered according to a stringent standardised quality control pipeline, as described previously (Lunnon et al., 2016) using the *WateRmelon* (Pidsley et al., 2013) package in R. We identified probes on the array within 1kb of differentially acetylated peaks (FDR < 0.05) using the Bioconductor package *GenomicRanges* (Lawrence et al., 2013). A total of 1,659 of the 4,162 FDR significant differentially acetylated peaks were located within 1kb of at least one CpG probe on the array, with a total of 6,838 probes mapping to the 1kb neighbourhood of these 1,649 peaks. For each CpG-peak pair we compared the log fold change in H3K27ac between AD cases and controls to the difference in DNA methylation or hydroxymethylation in a linear model controlling for the same covariates as in the differential acetylation analysis. Moreover, we compared DNA methylation and hydroxymethylation between probes in vicinity of AD hyper- and hypoacetylated peaks, as well as those in vicinity of all background peaks and the whole microarray background using t-tests. We used a hypergeometric test to characterize enrichment in intermediate methylation values (beta in [0.4,0.6]) in AD-dysregulated peaks compared to the array-wide background.

5.3 Results

5.3.1 Genome-wide profiling of inter-individual variation in H3K27ac in the entorhinal cortex

We generated H3K27ac ChIP-seq data using post-mortem entorhinal cortex tissue dissected from 47 elderly individuals (average age = 77.43, SD = 9.66, range = 58-97) comprising both AD cases (n = 24) and matched low pathology controls (n = 23) **(Table 5-1**). After stringent quality control (QC) of the raw data (see **Methods**), we obtained an average of 30,032,623 sequencing reads (SD = 10,638,091) per sample, with no difference in read-depth between cases and controls (*P* = 0.93, **Figure 5-2**). This represents, to our knowledge, the most extensive analysis of H3K27ac in the human entorhinal cortex yet undertaken. Using data from all 47 samples (see **Methods**) we identified 182,065 high confidence H3K27ac peaks; these are distributed across all 24 chromosomes (**Table 5-2**) spanning an average of 982bp (SD = 682bp) with a mean distance between neighbouring peaks of 15,536bp (SD = 116,040bp). We validated the identified peaks using two independent ChIP-seq datasets: first, we obtained locations for cortex (BA9) and cerebellum H3K27ac peaks from a recent publication by Sun and colleagues (Sun et al., 2016); second, we downloaded peak profiles for multiple cell- and tissue-types from the NIH Epigenomics Roadmap Consortium (Kundaje et al., 2015) (see **Methods**). There was a near perfect overlap between H3K27ac peaks called in other neocortical datasets and our ChIP-seq data, and a much lower overlap with H3K27ac data from non-cortical tissues (**Figure 5-5**). As a resource to the community, our data are available to download from GEO (accession number GSE102538).

Table 5-2. Distribution of H3K27ac peaks across the 24 chromosomes.

Chromosome	Number of peaks
1	17420
2	14739
3	11596
4	8136
5	10476
6	9783
7	9843
8	7921
9	8473
10	9064
11	9324
12	8670
13	4646
14	5927
15	6342
16	6972
17	8139
18	4008
19	5768
20	4885
21	2449
22	3873
X	3257
Y	354

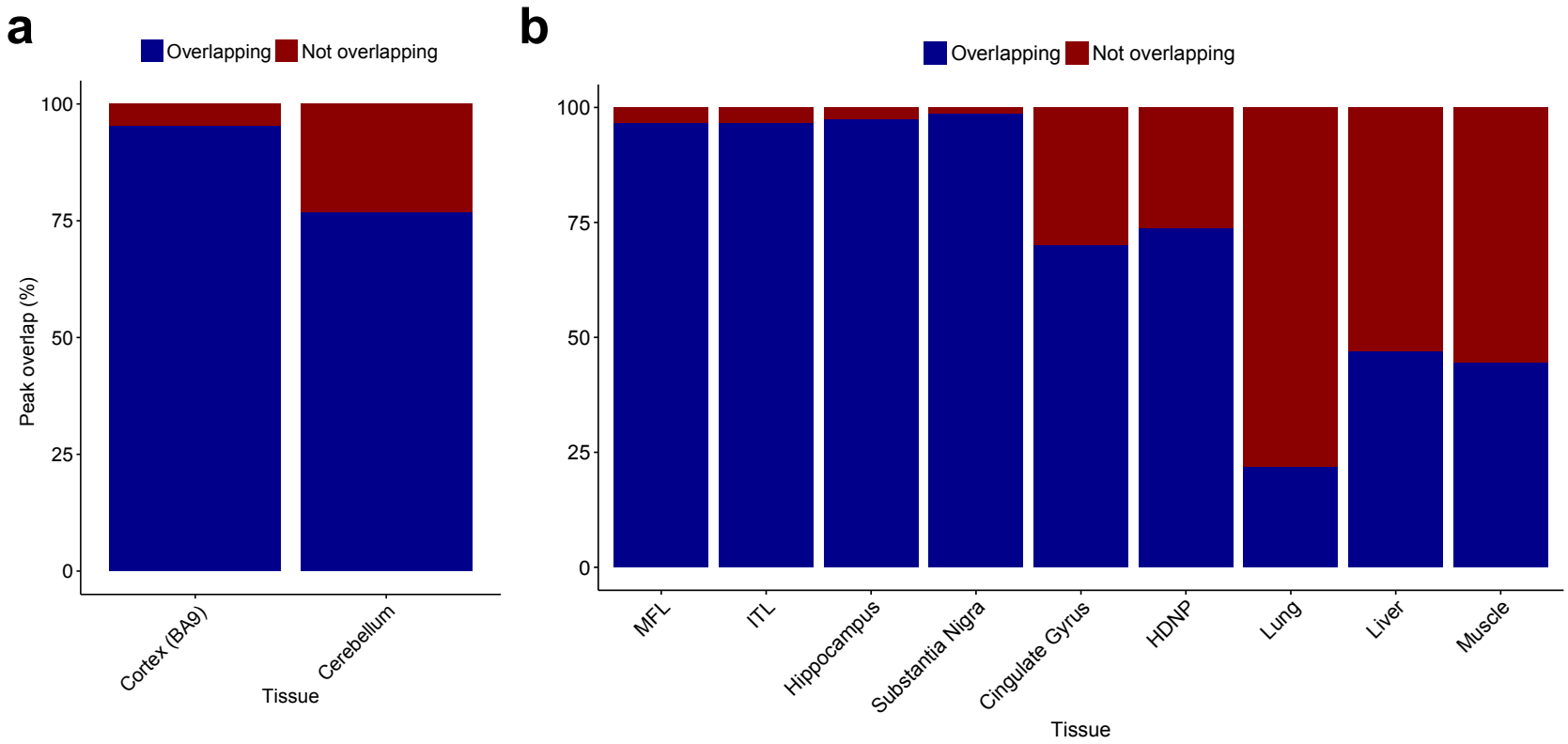


Figure 5-5. H3K27ac peaks identified in our study overlap substantially with those identified in other brain ChIP-seq datasets. Shown is the percentage of the peaks from other H3K27ac datasets overlapping with peaks in our entorhinal cortex (EC; n = 182,065) ChIP-seq data. (**a**) 53,882 (95%) of the 56,503 peaks identified by Sun et al (2016) in BA9 overlap our EC peaks, compared to 29,253 (77%) of the 38,069 cerebellum peaks. (**b**) Samples profiled by the Epigenomics Roadmap Consortium overlap our peaks in a tissue-specific manner with the highest overlap observed for cortical and sub-cortical datasets (mid frontal lobe (MFL): 97%, inferior temporal lobe (ITL): 97%, hippocampus: 97%, substantia nigra (99%), cingulate gyrus 70%, H1-derived neuronal progenitor cells (HDNP): 74%) and much lower overlaps for H3K27ac profiles derived from non-brain tissues (lung: 22%, liver: 47%, skeletal muscle: 44%).

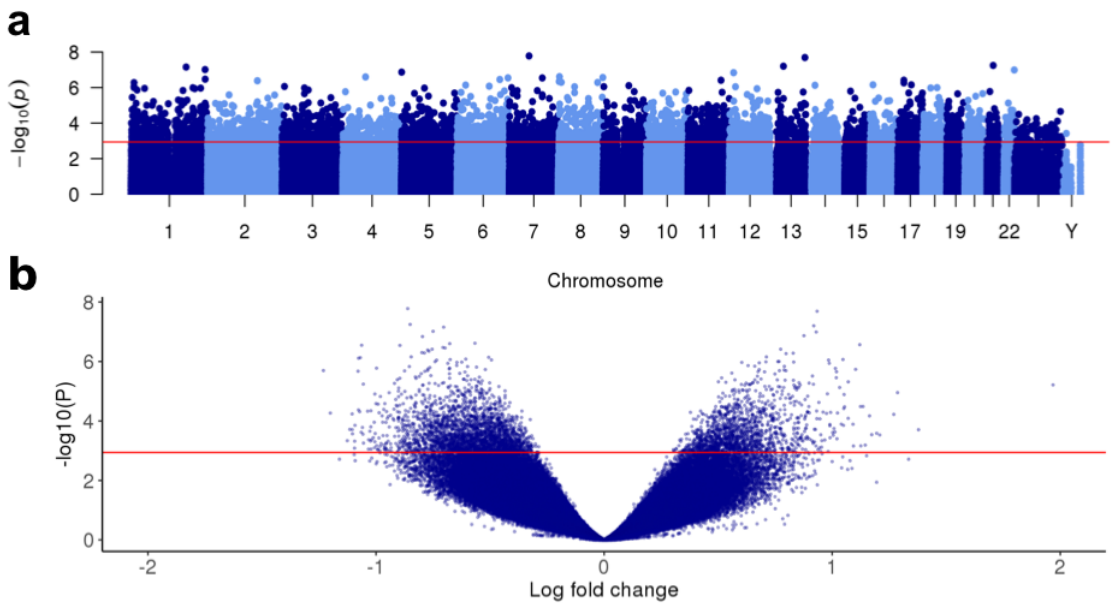


Figure 5-6. A genome-wide analysis of variable H3K27ac in AD identified widespread differentially acetylated peaks. (**a**) Manhattan plot showing the $-\log_{10}$ *P* value against chromosomal location from the *EdgeR* quasi-likelihood test, controlling for age and derived neuronal proportion. 4,162 peaks were identified as being associated with AD (FDR < 0.05, red line). (**b**) Volcano plot showing the $-\log_{10}$ *P* value against log fold change for each entorhinal cortex H3K27ac peak (FDR < 0.05, red line). Of the associated peaks, 1,475 (35%) are hyperacetylated in AD (higher H3K27ac) and 2,687 (65%) are hypoacetylated in AD (lower H3K27ac).

5.3.2 AD-associated differential acetylation in the entorhinal cortex

We quantified read counts for each peak in each individual sample using *HTSeq* (Anders et al., 2015) and employed the quasi-likelihood F test, implemented in the Bioconductor package *EdgeR* (Robinson et al., 2010), to test for differences in H3K27 acetylation associated with AD compared to low pathology controls (see **Methods**). Our analysis model controlled for age at death and neuronal cell proportion estimates derived from DNA methylation data generated on the same samples using the *CETS* R package (Guintivano et al., 2013) (**Table 5-1**, **Figure 5-4**). A total of 4,162 (2.3%) of the 182,065 peaks were characterized by differential acetylation associated with AD at a false discovery rate (FDR) < 0.05 (**Figure 5-6**), with a significant enrichment of hypo-acetylated AD-associated peaks (2,687 (1.5%)) compared to hyper-acetylated AD-associated peaks (1,475 (0.8%)) (*P* = 9.89E-80, exact binomial test) (**Figure 5-6**). The ten top-ranked hyper- and hypo-acetylated peaks associated with AD are shown in **Table 5-3,** with a full list given in **Appendix B-1** (hyperacetylated peaks) and **Appendix B-2** (hypoacetylated peaks). Peaks were subsequently annotated to genes using *GREAT* (McLean et al., 2010), which takes into account the strength of proximal

and distal DNA-binding events. The most significant AD-associated hyperacetylated peak (chr13: 112101248-112102698; *P* = 2.04E-08; log fold change = 0.93) is annotated to *SOX1* and *TEX29* on chromosome 13 (**Figure 5-7**, **Table 5-3**). H3K27ac data from the Epigenomics Roadmap Consortium show that this region is characterized by brain-specific enhancer activity (**Figure 5-8**). The most significant AD-associated hypoacetylated peak (chr7: 64011549-64012825; *P* = 1.66E-08; log fold change = -0.86) is located within intron 1 of *ZNF680* on chromosome 7 (**Figure 5-9**, **Figure 5-8**, **Table 5-3**). Global clustering of samples by normalized read counts across all hyper- and hypo-acetylated peaks (FDR < 0.05) indicated that samples group primarily by disease status (**Figure 5-10**).

AD-associated differentially acetylated peaks (FDR < 0.05) are significantly longer (1295bp vs 975bp, *P* < 1.00E-50) and characterized by higher read-depths (2.56 log counts per million (CPM) vs 1.59 log CPM, *P* < 1.00E-50) than non-significant peaks (**Figure 5-11**). Of note, within AD-associated peaks, hypoacetylated peaks are significantly longer (1412bp vs 1081bp, *P* = 5.66E-31) and have higher read depths (2.21 log CPM vs 1.77 log CPM, *P* = 2.69E-50) compared to hyperacetylated peaks. 43% of the overall peaks are located in regions classified as CpG islands, with differentially acetylated peaks significantly depleted in CpG islands (38%; *P* = 1.25E-05). However, amongst differentially acetylated peaks, 73% of hyperacetylated peaks are characterized as CpG islands compared to only 19% of hypoacetylated peaks (*P* < 1.0E-50; **Figure 5-11**). Conversely, differentially acetylated peaks are slightly enriched in regions mapping within 2kb upstream of protein coding genes (11%) compared to the background of non-significant peaks (9%; *P* = 5.13E-05), with no significant difference between hyper- and hypoacetylated peaks (**Figure 5-11**). Hyyperacetylated peaks show higher GC content (55%) than the background set of non-siginificant peaks (49%; *P* < 1.0E-50), while hypoacetylated peaks are depleted in GC content (44%; *P* < 1.0E-50) compared to the background (**Figure 5-11**). Finally, we tested for enrichments in CNVs (Zarrei et al., 2015) and STRs (Willems et al., 2014) using Fisher's exact test. While we found differentially acetylated peaks not to be enriched in CNVs (4% of differentially acetylated peaks overlap CNVs, compared to 6% in the non-significant background peak set), we observed an enrichment of STRs amongst AD-associated peaks (29%) compared to the non-significant background set of peaks (16%; *P* < 1.0E-50),

with STRs co-localizing with higher frequencies of CpG islands (47% CpG islands amongst STR peaks) compared to non-STR background peaks (42% CpG islands; $P < 1.0E-50$).

We used *RSAT* (Thomas-Chollier et al., 2012a, Thomas-Chollier et al., 2012b) to identify enriched transcription factor binding motifs located within AD-associated differentially acetylated peaks (see **Methods**), observing a significant enrichment of binding motifs for specificity protein 1 (Sp1) ($P < 1.0E-50$) amongst AD-hyperacetylated peaks (FDR < 0.05). Of note, previous publications have reported dysregulated expression of Sp1 and its co-localization with neurofibrillary tangles in AD (Santpere et al., 2006, Citron et al., 2008).

Table 5-3. Differential H3K27ac associated with AD. Shown are the ten top-ranked hyper- and hypo-acetylated H3K27ac peaks, controlling for age at death and neuronal proportion estimates derived from DNA methylation data.

Rank	CHR	Position (start – end)	*P* value	FDR	Log FC	Annotated genes
Hyperacetylated peaks						
1	13	112101248-112102698	2.04E-08	0.002	0.93	*SOX1, TEX29*
2	13	42094789-42095919	6.31E-08	0.003	0.92	*RGCC, VWA8*
3	22	50342521-50343567	1.02E-07	0.003	0.93	*PIM3, CRELD2*
4	5	640598-642071	1.36E-07	0.003	0.88	*CEP72, TPPP*
5	8	145180336-145181125	2.72E-07	0.004	1.12	*FAM203A, MAF1*
6	17	19665361-19666514	3.86E-07	0.004	0.77	*ALDH3A1, ULK2*
7	1	9392591-9393233	5.25E-07	0.004	0.83	*SLC25A33, SPSB1*
8	17	19619421-19620832	5.43E-07	0.004	0.80	*SLC47A2*
9	17	43925717-43927482	7.01E-07	0.005	0.71	*MAPT, SPPL2C*
10	1	9341867-9342320	8.55E-07	0.005	1.05	*SPSB1, H6PD*
Hypoacetylated peaks						
1	7	64011549-64012825	1.66E-08	0.002	-0.86	*ZNF680, ZNF736*
2	21	29827289-29828201	5.70E-08	0.003	-0.85	*N6AMT1*
3	1	179175226-179176637	7.03E-08	0.003	-0.70	*ABL2, TOR3A*
4	1	241397411-241399621	9.73E-08	0.003	-0.75	*GREM2, RGS7*
5	12	13627258-13629064	1.46E-07	0.003	-0.80	*EMP1, GRIN2B*
6	8	3964265-3966191	2.44E-07	0.004	-0.57	*CSMD1*
7	4	74088063-74089559	2.52E-07	0.004	-0.68	*COX18, ANKRD17*
8	6	166401119-166402753	2.85E-07	0.004	-1.06	*SDIM1, T*
9	7	107111795-107113029	2.88E-07	0.004	-0.90	*DUS4L, GPR22*
10	1	241694436-241695782	3.37E-07	0.004	-0.71	*KMO*

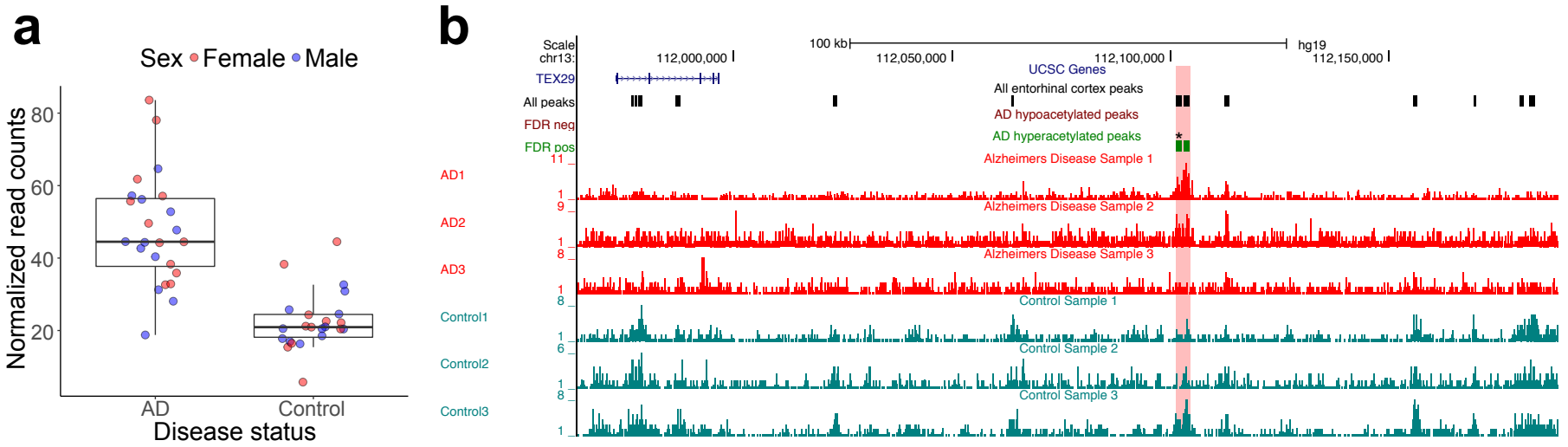
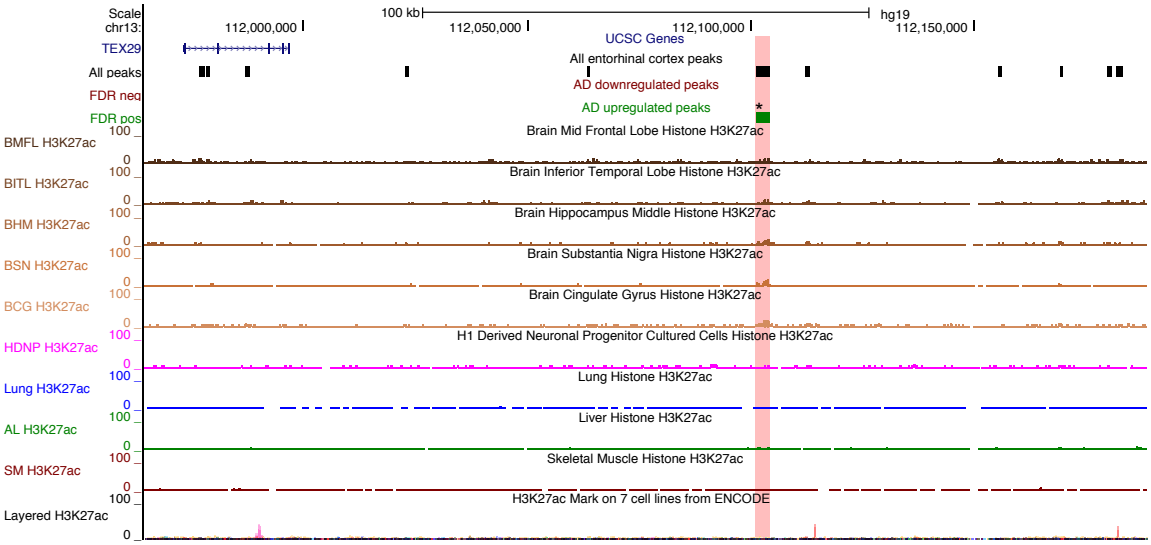


Figure 5-7. The top-ranked AD-associated hyperacetylated peak is annotated to *SOX1* and *TEX29* on chromosome 13. Shown are normalized read counts (**a**) and a regional track of H3K27ac ChIP-seq profiles from three example cases (red) and controls (turquoise; **b**). (**a**) The most significant AD-hyperacetylated peak is characterized by a consistent increase of H3K27ac in diseased samples (*P* = 2.04E-08, log fold change = 0.93). (**b**) This peak is located on chromosome 13 and annotated to *SOX1* and *TEX29*.

a



b

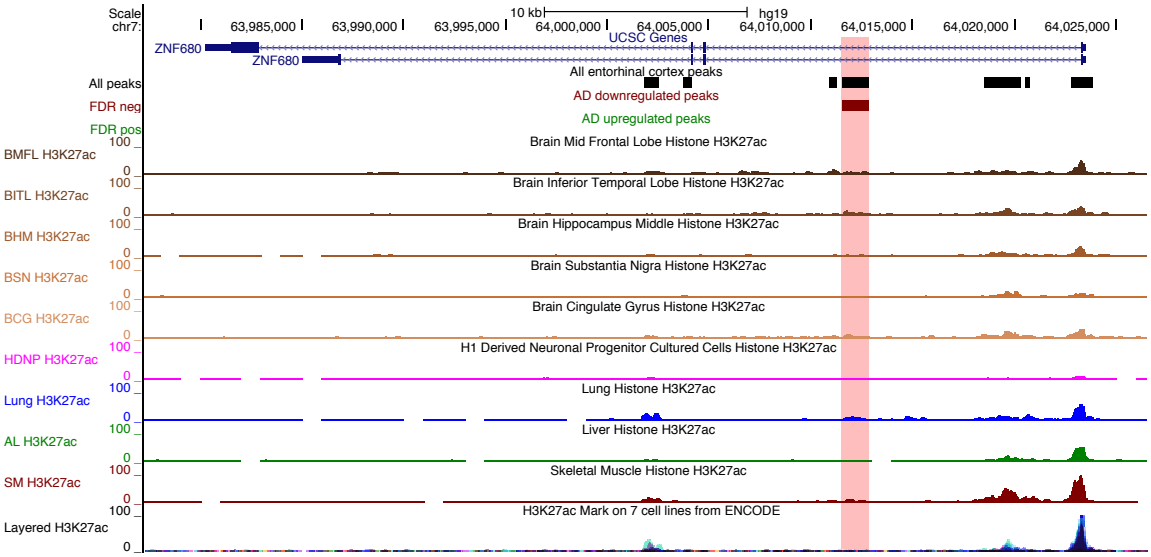


Figure 5-8. Regional profiles of H3K27ac signals in multiple cell- and tissue types around the top-ranked AD-associated differentially acetylated peaks. H3K27ac ChIP-seq profiles for several NIH Epigenomics Roadmap Consortium samples are shown: mid frontal lobe (BMFL), inferior temporal lobe (BITL), middle hippocampus (BHM), substantia nigra (BSN), cingulate gyrus (BCG), H1-derived neuronal progenitor cells (HDNP), lung, liver (AL), skeletal muscle (SM) and the layered H3K27ac track based on multiple cell lines displayed by the UCSC Genome Browser in the default setting. (**a**) A region surrounding the most significantly upregulated peak (*P* = 2.04E-08; highlighted by an asterix) on chromosome 13 is presented. The peak and its immediate neighbour are characterized by brain-specific H3K27ac profiles. (**b**) A region around the most significantly downregulated peak (*P* = 1.66E-08), located in an intron of *ZNF680* on chromosome 7, is shown.

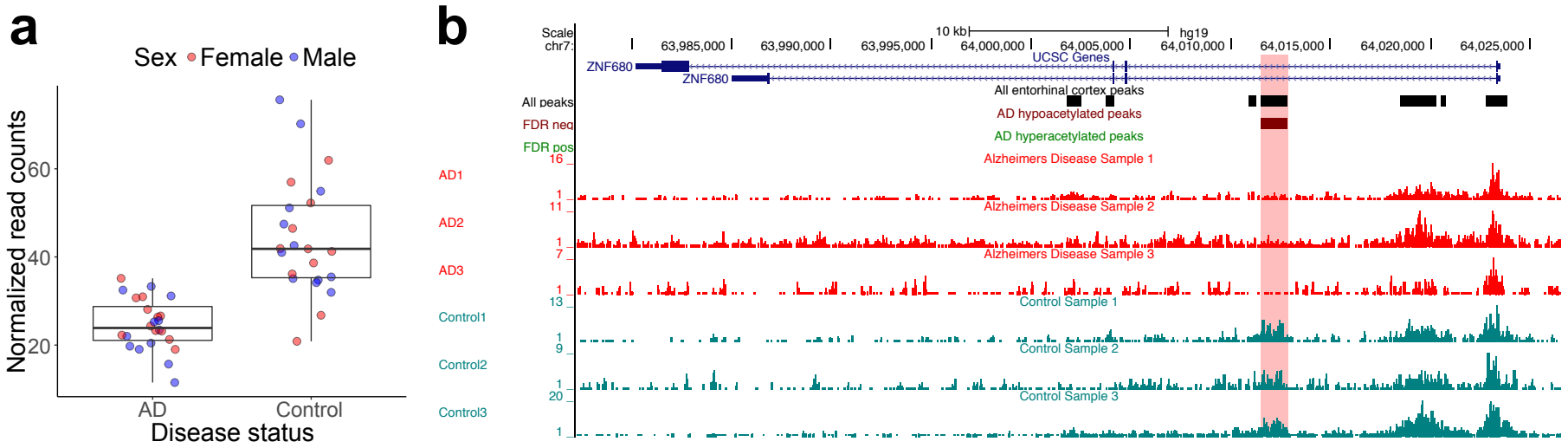


Figure 5-9. The top-ranked AD-associated hypoacetylated peak is located in intron 1 of *ZNF680* on chromosome 7. Shown are normalized read counts (**a**) and regional tracks of H3K27ac ChIP-seq profiles from three example cases (red) and controls (turquoise; **b**). (**a**) The most significant differentially acetylated peak (*P* = 1.66E-08) exhibits decreased H3K27ac in AD cases (log fold change = -0.86). (**b**) This AD-hypoacetylated peak is located in intron 1 of *ZNF680* on chromosome 7.

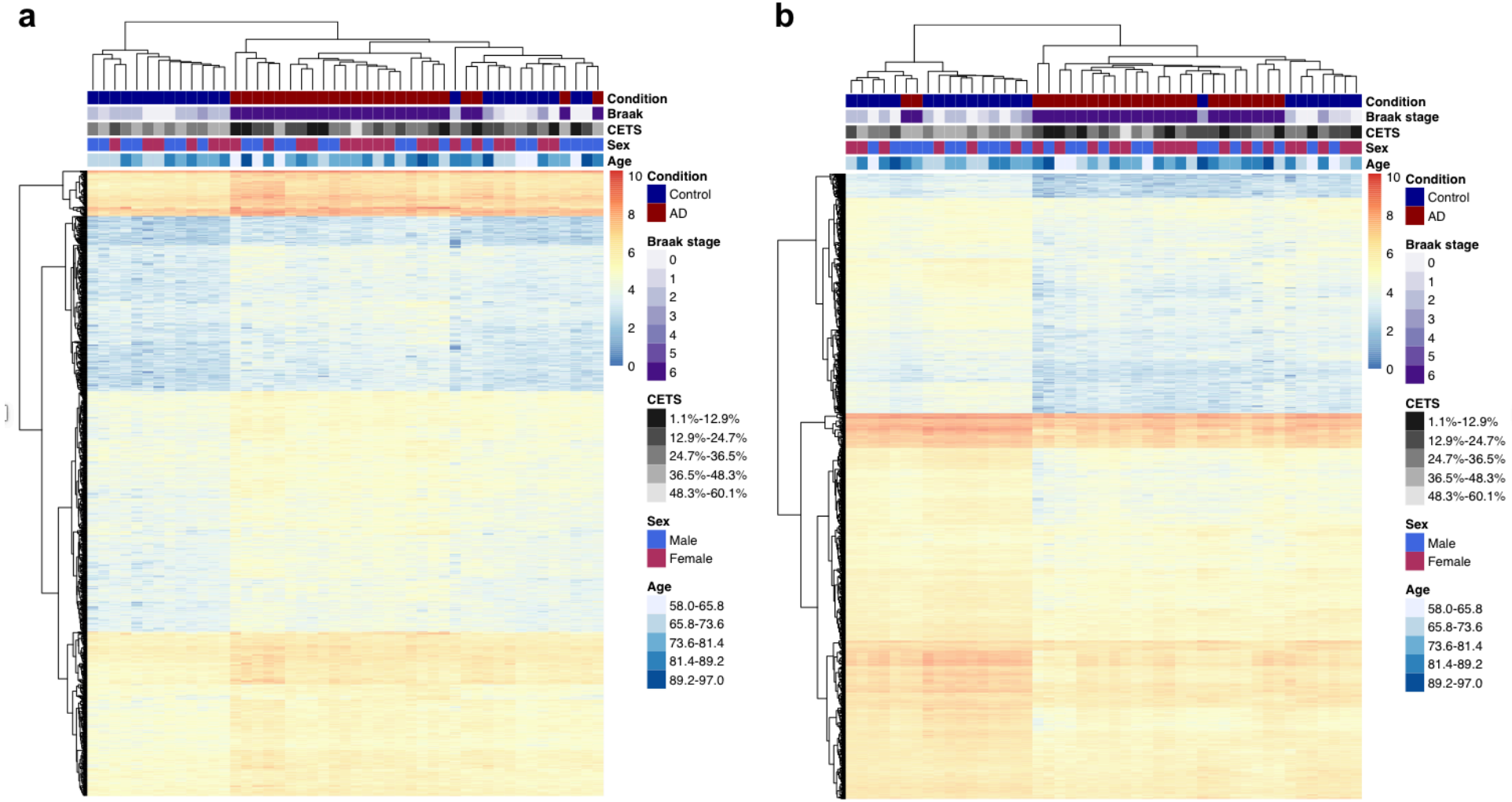

a
Condition
Braak
CETS
Sex
Age
10
8
6
4
2
0
Condition
Control
AD
Braak stage
0
1
2
3
4
5
6
CETS
1.1%-12.9%
12.9%-24.7%
24.7%-36.5%
36.5%-48.3%
48.3%-60.1%
Sex
Male
Female
Age
58.0-65.8
65.8-73.6
73.6-81.4
81.4-89.2
89.2-97.0
b
Condition
Braak stage
CETS
Sex
Age
10
8
6
4
2
0
Condition
Control
AD
Braak stage
0
1
2
3
4
5
6
CETS
1.1%-12.9%
12.9%-24.7%
24.7%-36.5%
36.5%-48.3%
48.3%-60.1%
Sex
Male
Female
Age
58.0-65.8
65.8-73.6
73.6-81.4
81.4-89.2
89.2-97.0

Figure 5-10. Clustering of samples by H3K27ac at differentially acetylated peaks groups them primarily by disease status. Shown are two heatmaps clustering samples based on the 1,475 hyperacetylated peaks (**a**) and the 2,687 hypoacetylated peaks (**b**), respectively. (**a**) A heatmap, clustering samples by normalized read counts in all significant AD hyperacetylated peaks (FDR < 0.05), generates three distinct groups: one comprised of controls only (group 1, 13 samples, left), a pure group of cases (group 2, 20 samples, middle), and a mixed group containing both cases and controls (group 3, 14 samples, right). The differences in the disease status distribution across these three groups is significant (*P* = 1.89E-08, chi-square test with Yates' continuity correction). Controls grouped together with cases in group 3 are characterized by significantly decreased neuronal proportion estimates, compared to those in the pure control group 1 (*P* = 7.10E-04, t-test). (**b**) A heatmap, clustering samples by all FDR significant AD hypoacetylated peaks, divides the samples into two main groups: group 1 is composed mainly of controls (17 samples, left), whereas group 2 contains more cases than controls (30 samples, right). Case-control differences between these two groups are significant (*P* = 1.74E-04, chi-square test). Interestingly, controls classified into group 2 are characterized by lower neuronal proportion estimates than those in group 1 (*P* = 0.004, t-test). The clustering defined by hyper- or hypoacetylated peaks is not significantly associated with sex (*P* > 0.05, chi-square test) or age at death (*P* > 0.05, t-test).

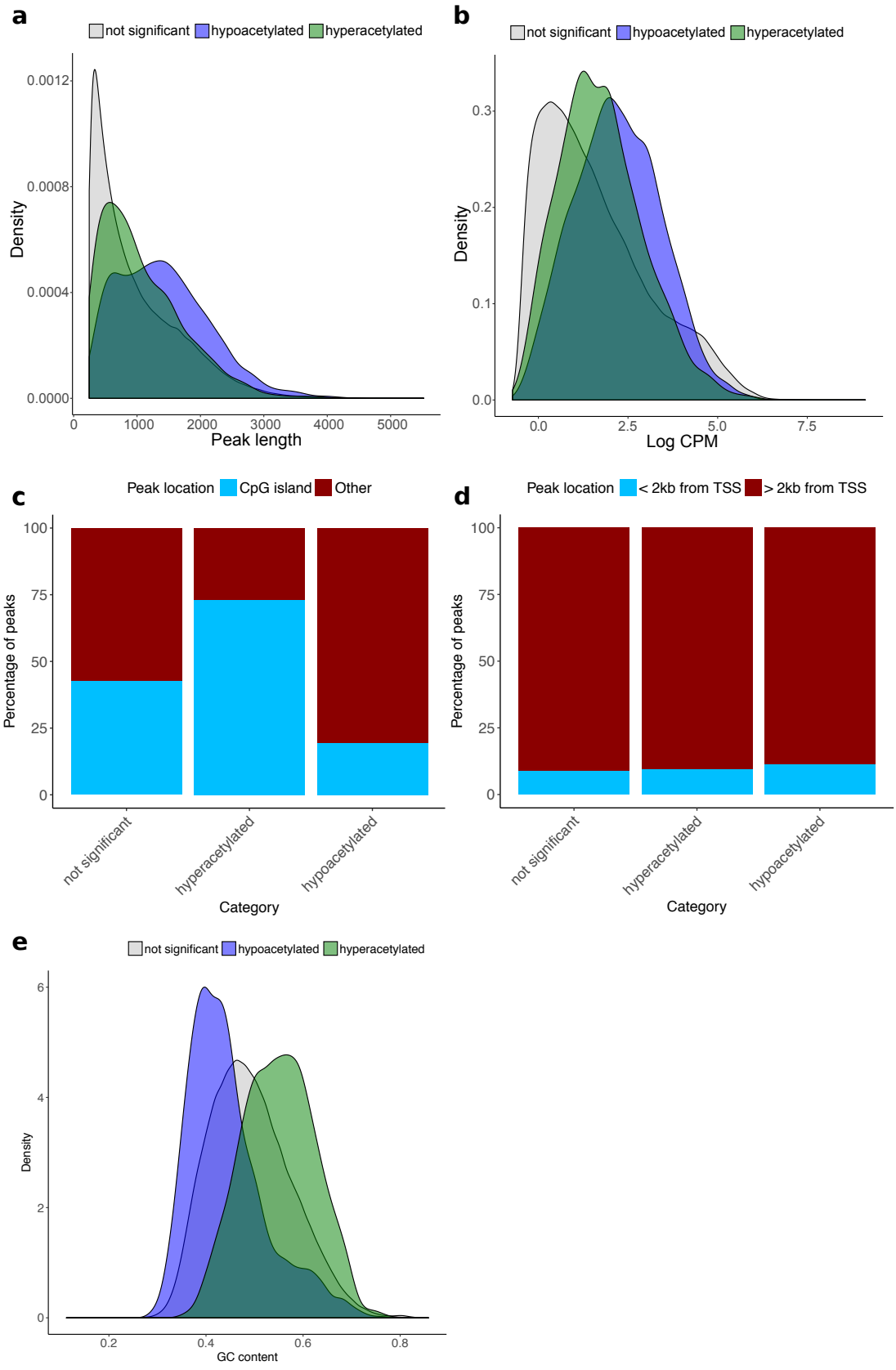


Figure 5-11. Peak characteristics of differentially acetylated peaks. On average, FDR significant differentially acetylated peaks (**a**) are longer (P < 1.00E-50) and (**b**) have higher log counts per million (CPM; P < 1.00E-50) compared to the non-significant background peaks. Interestingly, amongst differentially acetylated peaks, those that are hypoacetylated are characterized by (**a**) being longer (P = 5.66E-31) and (**b**) having higher CPM (P = 2.69E-50) than the significantly hyperacetylated peaks. (**c**) Overall, differentially acetylated peaks are significantly depleted in CpG island peaks compared to non-significant background peaks (P =

1.25E-08). However, amongst differentially acetylated peaks, hyperacetylated peaks show a larger percentage of CpG island peaks compared to hypoacetylated peaks (P < 1.0E-50). (**d**) In contrast, differentially acetylated peaks are enriched in regions mapping within 2kb upstream of the TSS of protein coding genes (P = 5.13E-05) but there is no statistically significant difference between hyper- and hypoacetylated peaks (P = 0.08). (**e**) Compared to the background GC density (49%), hyperacteylated peaks are enriched in GC content (55%, P < 1.0E-50) and hypoacetylated peaks are depleted in GC content (44%, P < 1.0E-50).

5.3.3 Increased H3K27ac is observed in regulatory regions annotated to genes previously implicated in both tau and amyloid neuropathology

One of the top-ranked AD-associated hyper-acetylated peaks is located proximal to the gene encoding microtubule associated protein tau (MAPT) (chr17: 43925717-43927482; P = 7.01E-07; log fold change = 0.71; **Table 5-3**), which is widely expressed in the nervous system acting to promote microtubule assembly and stability. Tau is believed to play a key role in AD neuropathology, with hyperphosphorilation of the tau protein precipitating the neurofibrillary tangles associated with the pathogenesis of AD (Ittner and Gotz, 2011, Spillantini and Goedert, 2013). Closer inspection of the region around this AD-associated peak highlighted an extended cluster of six hyper-acetylated H3K27ac peaks (FDR < 0.05) spanning 36 kb (chr17: 43925717 - 43961546) located within a *MAPT* antisense transcript (*MAPT_AS1*) ~10kb upstream of the *MAPT* transcription start site (**Figure 5-12**; **Table 5-4**). H3K27ac ChIP-seq data from the NIH Epigenomics Roadmap Consortium show that this region is characterized by brain-specific H3K27ac signatures (**Figure 5-12**), with *ChromHMM* (Ernst and Kellis, 2012) identifying the region as an active chromatin domain in brain, comprised of enhancers and blocks of weak transcription (**Figure 5-13**). Strikingly, AD-associated differentially-acetylated peaks were also found in the vicinity of other genes known to play a direct mechanistic role in AD. We identified a significantly hypoacetylated peak (chr21:27160993-27161475; P = 3.94E-04; log fold change = -0.72) on chromosome 21, located ~100 kb downstream of the amyloid precursor protein gene (*APP*), which encodes the precursor molecule to Aβ, the main component of amyloid plaques (Rovelet-Lecrux et al., 2006, Scheuner et al., 1996, Goate et al., 1991, Jonsson et al., 2012, Cruchaga et al., 2012) (**Figure 5-14**). We also identified significant hyperacetylation in the vicinity of the presenilin genes *PSEN1* and *PSEN2*, which encode integral components of the gamma secretase complex and play a key role in generation of Aβ from APP (De Strooper et al., 1998). In *PSEN1* we found significantly elevated

H3K27ac across a peak within intron 6 (chr14:73656445-73656860; *P* = 3.44E-04; log fold change = 0.68; **Figure 5-15**). In *PSEN2* we identified consistent hyperacetylation in AD cases across nine H3K27ac peaks (FDR < 0.05) spanning a ~57 kb region upstream of the transcription start-site (chr1:226957425-227014019; **Figure 5-16, Figure 5-13** and **Table 5-5**). Of note, highly-penetrant mutations in *APP*, *PSEN1*, and *PSEN2* are associated with familial forms AD (Goate and Hardy, 2012, Tanzi, 2012, Schellenberg and Montine, 2012). The identification of altered regulation of these loci in late-onset sporadic AD brain further implicates a role for altered amyloid processing in the onset of neuropathology.

Table 5-4. Association statistics for the nine peaks overlapping the differentially acetylated region on chromosome 17, annotated to *MAPT* and *SPPL2C.*

Start - end	*P*-value	*P* FDR	Log FC
43925717-43927482	7.01E-07	0.005	0.71
43928851-43929350	7.68E-04	0.043	0.60
43929961-43930229	1.08E-03	0.049	0.59
43935330-43936111	1.72E-04	0.024	0.54
43938147-43939070	0.842	0.941	0.02
43943325-43946054	0.165	0.454	0.20
43947431-43948194	8.84E-04	0.046	0.45
43954021-43954371	0.052	0.263	0.46
43959954-43961546	8.94E-05	0.020	0.59

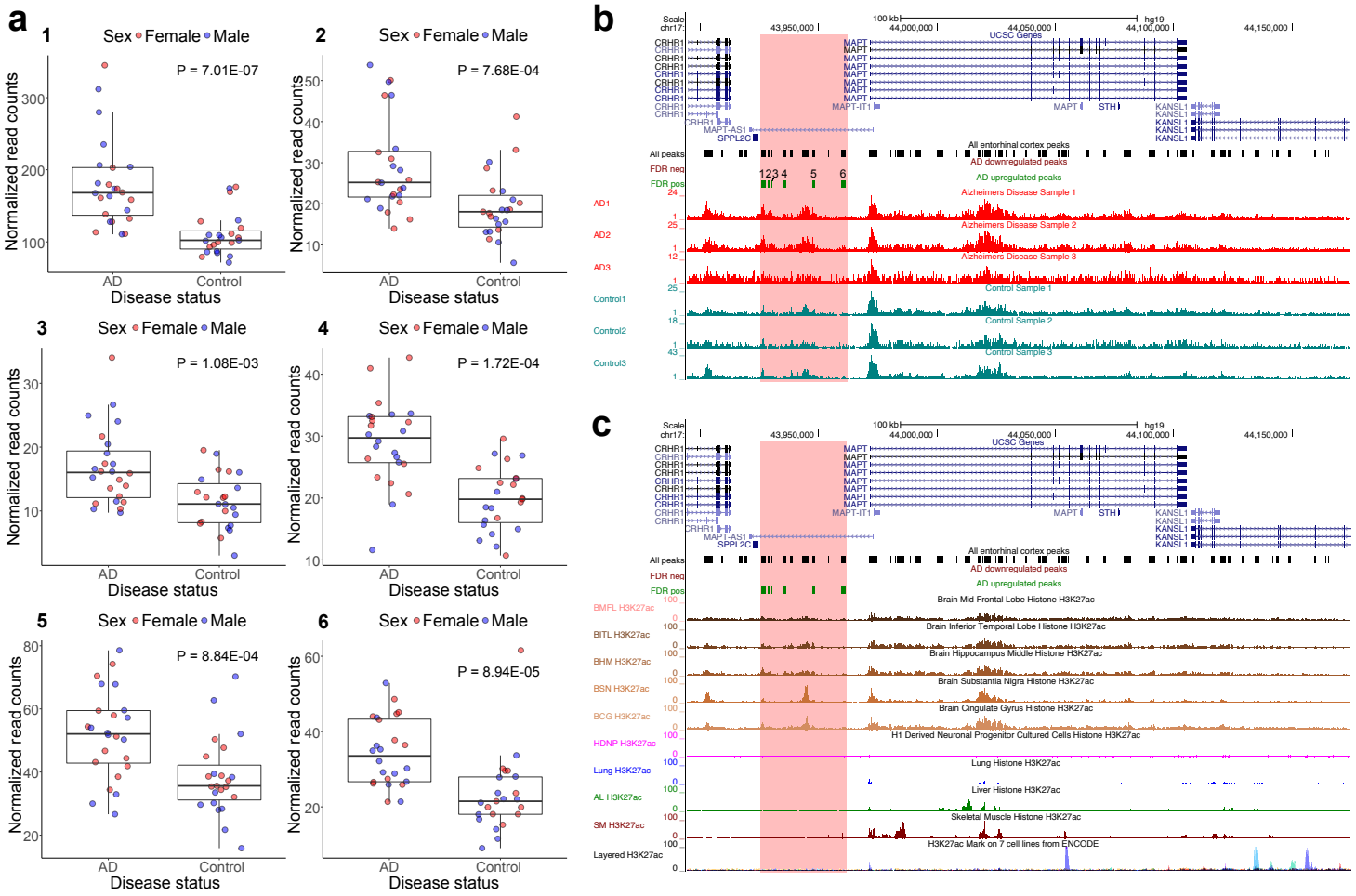


Figure 5-12. A region annotated to *MAPT* comprising six H3K27ac peaks is characterized by significant hyperacetylation in AD. A cluster of nine H3K27ac peaks was identified on chromosome 17. All nine peaks show consistent effect directions (mean log fold change = 0.46; **Table 5-4**), with H3K27ac upregulated in cases. (**a**) For six of the nine peaks this increase in H3K27ac associated with AD is significant (FDR < 0.05). (**b**) The whole region is located 10kb upstream of *MAPT* (**c**) and is characterized by brain specific acetylation profiles. The boundaries of the significantly differentially acetylated peak region are highlighted in red.

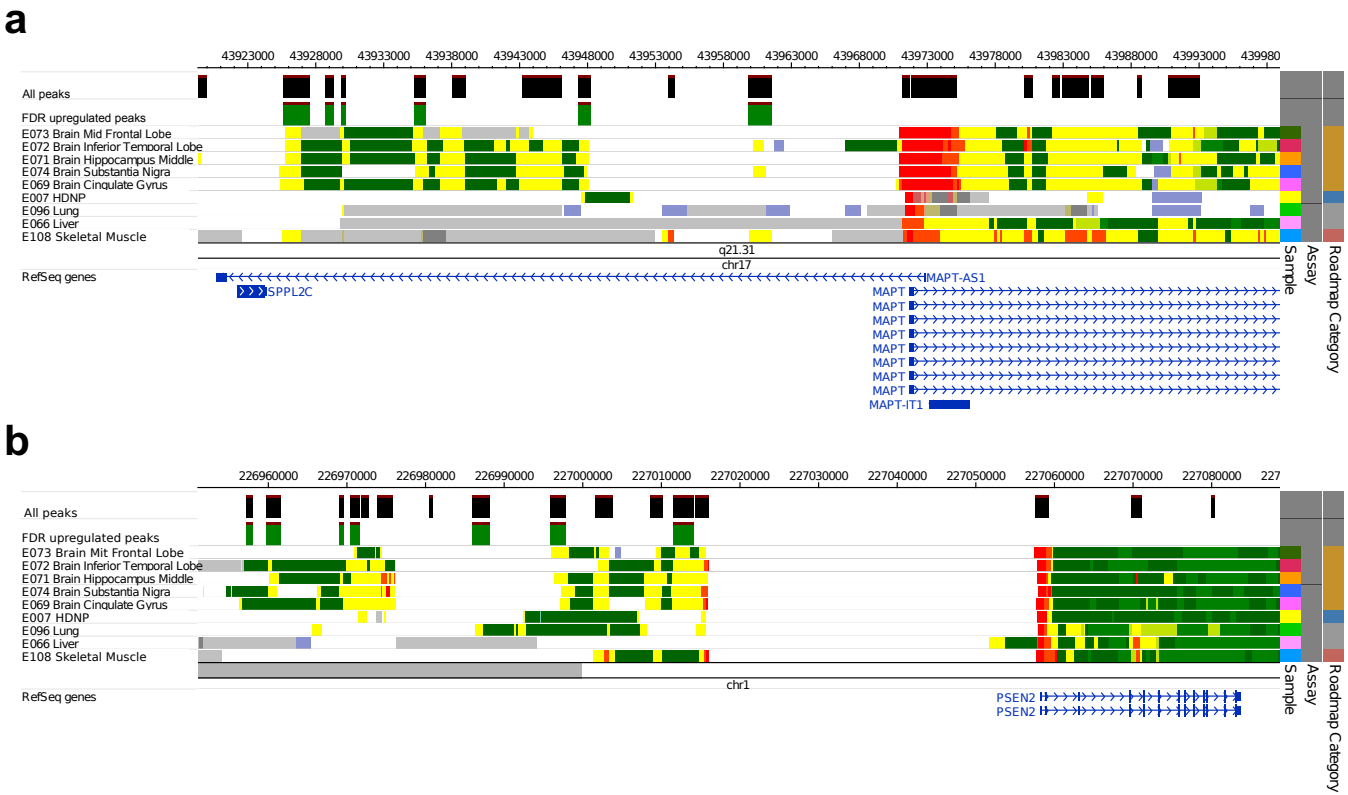


Figure 5-13. Chromatin state tracks for the *MAPT* and *PSEN2* hyperacetylated peak clusters. The tracks show the locations of H3K27ac peaks across the plotted window (black) as well as the peaks showing significant increase of H3K27ac in AD (FDR < 0.05; green). ChromHMM chromatin states from the Roadmap Epigenomics Consortium for a range of brain tissues (mid frontal lobe, inferior temporal gyrus, hippocampus middle, substantia nigra and cingulate gyrus) as well as non-brain tissue (H1-derived neuronal progenitor cells, lung, liver and skeletal muscle). Plots were produced on the WashU Epigenome Browser v42 using the core 15-state ChromHMM model. Chromatin states matching the colours shown include red - active TSS, green – strong transcription, dark green – weak transcription, yellow – enhancer, gray – repressed polycomb, light gray – weak repressed polycomb, pale turquoise – heterochromatin, and white – quiescent. (**a**) A cluster of six AD-hyperacetylated H3K27ac peaks was identified upstream of *MAPT*. Chromatin state annotation shows that this region is characterized primarily by brain-specific enhancers and blocks of weak transcription. (**b**) A second cluster spanning nine significantly hyperacetylated peaks was identified upstream of *PSEN2*. ChromHMM identified two active chromatin domains in brain, also comprised of enhancers and blocks of weak transcription.

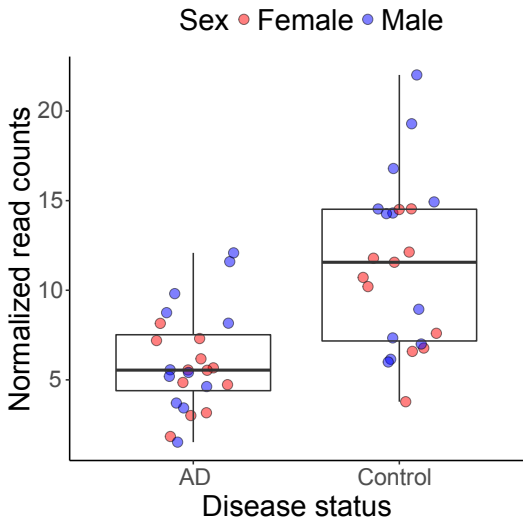


Figure 5-14. A H3K27ac peak located downstream of *APP* is characterized by significant hypoacetylation in AD. We identified one significantly hypoacetylated peak (FDR < 0.05) on chromosome 21, located < 100 kb downstream of *APP* (chr21:27160993-27161475; *P* = 3.94E-04; log fold change = -0.72).

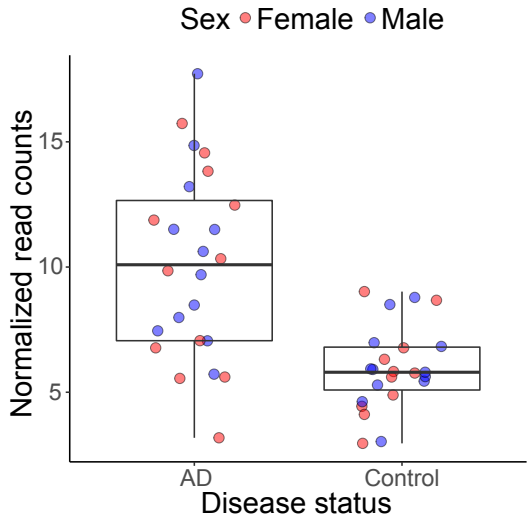


Figure 5-15. A H3K27ac peak located in an intron of *PSEN1* is characterized by significant hyperacetylation in AD. We identified one significantly hyperacetylated peak (FDR < 0.05) on chromosome 14, located in intron 6 of *PSEN1* (chr14:73656445-73656860; *P* = 3.44E-04; log fold change = 0.68).

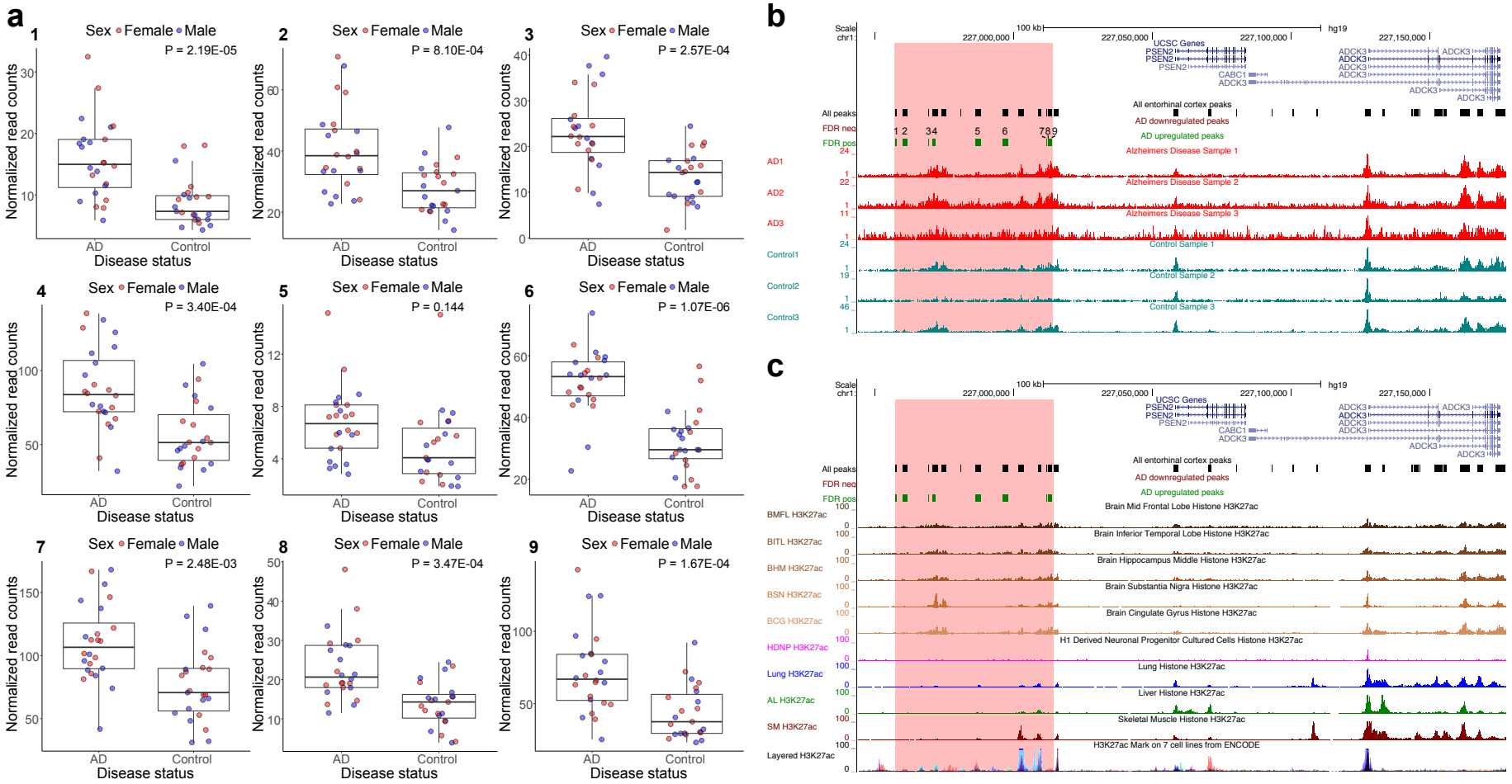


Figure 5-16. A region annotated to *PSEN2* spanning nine H3K27ac peaks is characterized by significant hyperacetylation in AD. A cluster of 14 H3K27ac peaks was identified on chromosome 1. All 14 peaks show consistent effect directions (mean log fold change = 0.52; Supplementary Table 6), with H3K27ac upregulated in cases. (**a**) For nine of the 14 peaks this increase in H3K27ac associated with AD is significant (FDR < 0.05). (**b**) The whole region is located 44kb upstream of *PSEN2* (**c**) and is characterized by predominantly acetylation peaks in brain. The boundaries of the significantly differentially acetylated peak region are highlighted in red.

Table 5-5. Association statistics for the 14 peaks overlapping the differentially acetylated region on chromosome 1, annotated to *PSEN2* and *ITPKB.*

Start - end	*P*-value	*P* FDR	Log FC
226957424-226958029	2.19E-05	0.013	0.84
226959991-226961572	8.10E-04	0.044	0.46
226969195-226969518	2.57E-04	0.029	0.66
226970654-226971677	3.40E-04	0.032	0.56
226971888-226972730	0.014	0.144	0.39
226974106-226975826	0.014	0.144	0.37
226980648-226980894	0.144	0.426	0.34
226986193-226988116	1.07E-06	0.005	0.60
226996028-226997864	6.10E-05	0.018	0.53
227001743-227003902	0.012	0.134	0.28
227008727-227010081	2.48E-03	0.068	0.42
227011597-227011851	3.47E-04	0.032	0.60
227012069-227012797	1.67E-04	0.024	0.65
227013015-227014019	7.90E-04	0.043	0.53

5.3.4 Specific differentially-acetylated peaks also overlap known AD GWAS regions

Using data from a large genome-wide association study (GWAS) of AD (Lambert et al., 2013) we tested for instances where there is an overlap between AD-associated differential H3K27ac and genomic regions identified in GWAS. Briefly, we defined linkage-disequilibrium (LD) blocks around the 11 GWAS variants identified by Lambert and colleagues (**Table 5-6**) which contained a total of 292 overlapping entorhinal cortex H2K27ac peaks (see **Methods**). Two of the 11 GWAS LD blocks contained significant AD-associated H3K27ac peaks (FDR < 0.05), although there was no overall enrichment of AD-associated differential acetylation at the 292 peaks (P = 0.364, Wilcoxon rank-sum test). Two peaks of AD-associated hyperacetylation were located within a GWAS region on chromosome 1, mapping to the gene body of *CR1* (chr1: 207753457-207753813; P = 1.15E-06; log fold change = 0.99 and chr1: 207754916-207756572; P = 5.40E-04; log fold change = 0.56; **Figure 5-17**). *CR1* encodes a transmembrane glycoprotein expressed in microglia with a role in the innate immune system, promoting phagocytosis of immune complexes and cellular debris, in addition to Aβ (Crehan et al., 2012, Heppner et al., 2015, Villegas-Llerena et al., 2016). Two

other AD-associated differentially acetylated peaks were found to be located within a GWAS region on chromosome 19, including a hyperacetylated peak (chr19:45394441- 45395396; *P* = 2.13E-04; log fold change = 0.48) mapping to the gene body of *TOMM40* in the immediate vicinity of *APOE* (**Figure 5-18**). Another H3K27ac peak in this LD block was hypoacetylated in AD (chr19: 45639588- 45641733; *P* = 7.65E-04; log fold change = -0.33), mapping to intron 1 of *PPP1R37*.

Table 5-6. Shown are the locations, GWAS SNPs and *P* values for the 11 LD blocks constructed from the Lambert et al AD GWAS.

CHR	Start	End	Number of SNPs in GWAS	Min P SNP	Location	P
1	207372941	207875537	1144	rs1752684	207747296	3.65E-15
2	127826533	127889932	255	rs7561528	127889637	6.54E-18
6	47318885	47662484	981	rs9381563	47432637	5.30E-09
7	143083661	143198249	278	rs10808026	143099133	1.42E-11
8	27170788	27489959	869	rs7982	27462481	2.48E-17
11	59826677	60105199	733	rs72924659	60103385	5.35E-13
11	85636074	85874322	648	rs10792832	85867875	6.53E-16
11	121423640	121517613	176	rs11218343	121435587	4.98E-11
14	92926952	92941096	95	rs12590654	92938855	4.10E-08
18	29088958	29088958	1	rs8093731	29088958	4.63E-08
19	45050747	45691126	1504	rs12972156	45387459	<1.00E-50

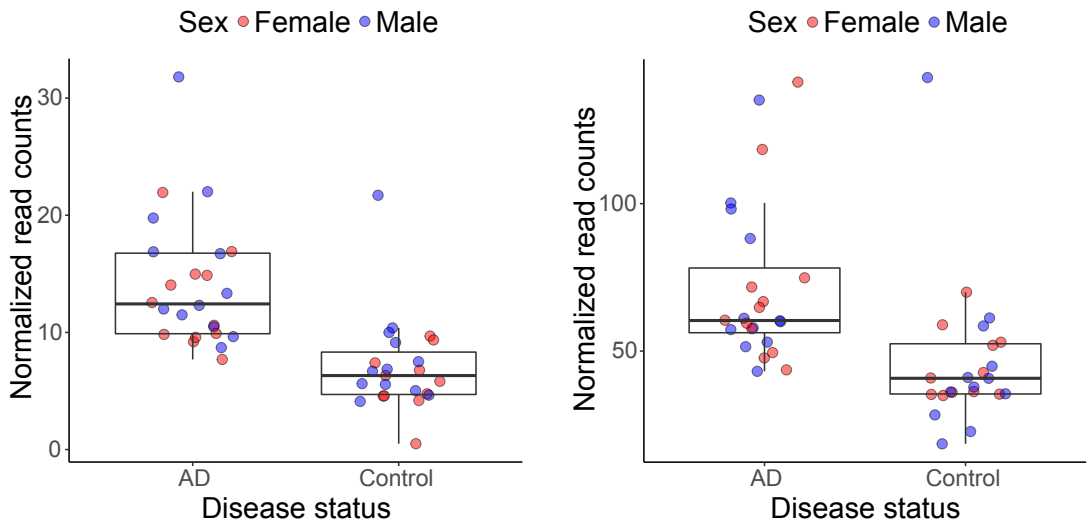


Figure 5-17. Two intragenic H3K27ac peaks located across *CR1* are characterized by significant hyperacetylation in AD. We identified two significantly hyperacetylated peak (FDR < 0.05) on chromosome 1, located within the gene body of *CR1*: (**a**) chr1: 207753457-207753813;

P = 1.15E-06; log fold change = 0.99 and (**b**) chr1: 207754916-207756572; *P* = 5.40E-04; log fold change = 0.56.

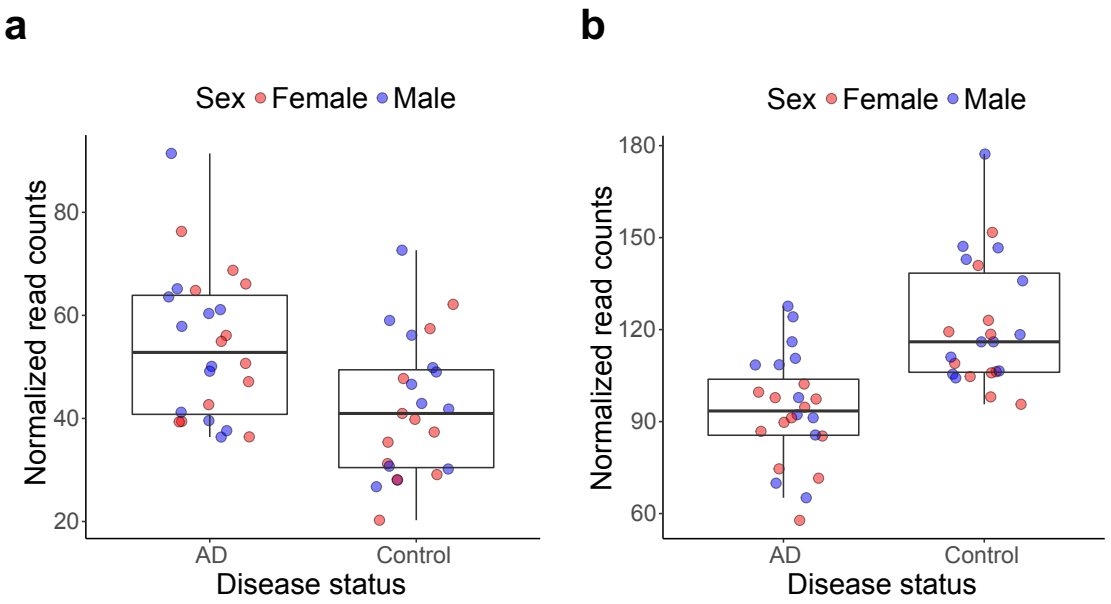


Figure 5-18. Two differentially acetylated peaks overlap the GWAS LD block on chromosome 19. (**a**) We identified one significantly hyperacetylated peak (FDR < 0.05) on chromosome 19, mapping to the gene body of *TOMM49* and < 15 kb upstream of *APOE* (chr19:45394441- 45395396; *P* = 2.13E-04; log fold change = 0.48). (**b**) A peak located in intron 1 of *PPP1R37* was hypoacetylated in AD (chr19: 45639588- 45641733; *P* = 7.65E-04; log fold change = -0.33).

5.3.5 AD-associated differentially-acetylated peaks are enriched for functional processes related to neuropathology

We used *GREAT* (McLean et al., 2010) to calculate statistical enrichments for ontological annotations, interrogating gene ontologies for molecular function and biological processes (Ashburner et al., 2000), human diseases (Osborne et al., 2009) as well as human phenotypes (Robinson and Mundlos, 2010) (**Figure 5-19** and **Appendix C**). Multiple ontological categories associated with AD progression and pathology were identified as being enriched amongst hyperacetylated peaks, including "lipoprotein particle binding" (Jaeger and Pietrzik, 2008, Liu et al., 2013) (*P* = 1.10E-06), "beta-amyloid metabolic process" (Ittner and Gotz, 2011) (*P* = 4.94E-08), "neurofibrillary tangles" (Spillantini and Goedert, 2013) (*P* = 3.93E-05), "response to hypoxia" (Sun et al., 2006, Zlokovic, 2011) (*P* = 3.17E-14), "neuronal loss in central nervous system" (Wenk, 2003) (*P* = 4.93E-04), and "Pick's disease" (*P* = 2.93E-07), a form of fronto-temporal dementia also characterized by tau pathology (Warren et al., 2013, Spillantini and Goedert, 2013). Amongst hypoacetylated peaks we observed an enrichment of

categories related to neurotransmitter-functions, including “GABA receptor activity” (Limon et al., 2012) (*P* = 2.70E-07) as well as categories related to neuronal transmission and synapses, such as “protein location to synapse” (*P* = 7.86E-09).

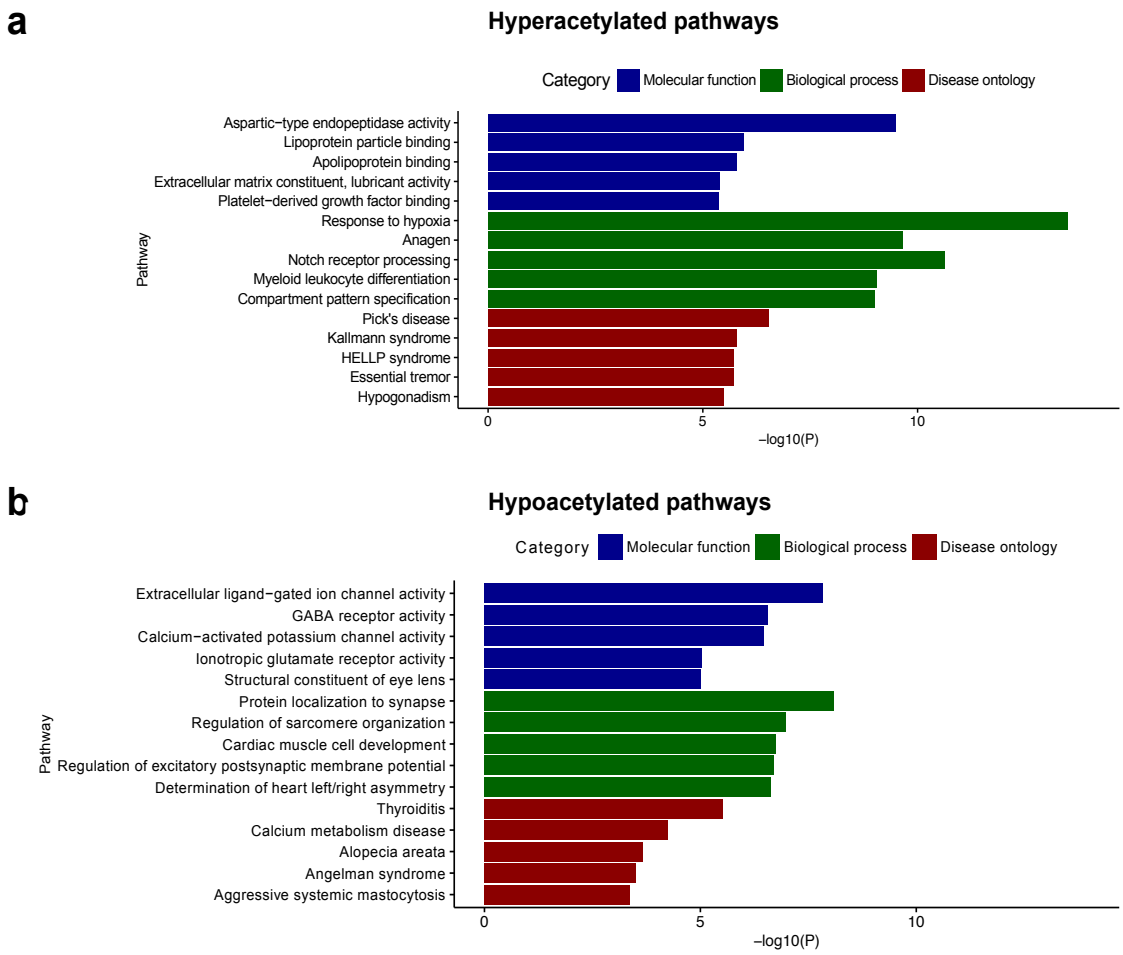


Figure 5-19. Neurobiological and disease-related pathways are enriched in hyper- and hypoacetylated regions. Using all significantly hyper- and hypoacetylated peaks (FDR <0.05), we conducted functional enrichment analyses using *GREAT* (McLean et al., 2010). Shown are the top five independent enrichments in the categories molecular function, biological process and disease ontology (see **Methods**). (**a**) AD-related pathways enriched in the hyperacetylated peaks include “lipoprotein particle binding” (*P* = 1.10E-06) and “response to hypoxia” (*P* = 3.17E-14) as well as “Pick’s disease” (*P* = 2.93E-07), a form of fronto-temporal dementia. (**b**) Amongst hypoacetylated pathways we identified neuronal transmission pathways, including “protein location to synapse” (*P* = 7.86E-09) and “GABA receptor activity” (*P* = 2.70E-07).

5.3.6 Integrative analysis of DNA and histone modifications reveals unique distributions of DNA modifications across regions of differential acetylation

Our previous work identified cortex-specific variation in DNA methylation (5mC) associated with AD pathology (Lunnon et al., 2014, De Jager et al., 2014). We were therefore interested in exploring the relationship between H3K27ac and both 5mC and another DNA modification – DNA hydroxymethylation (5hmC), which is enriched in the brain and believed to play an important role in neuronal function, learning and memory (Lunnon et al., 2016, Kaas et al., 2013, Kinde et al., 2015) - in AD. Both modifications were profiled using DNA isolated from the same entorhinal cortex samples using oxidative bisulfite conversion in conjunction with the Illumina 450K HumanMethylation array ("450K array") (see **Methods**). 6,838 probes mapped to within 1kb of differentially acetylated peaks (FDR < 0.5; mapping to 1,649 unique peaks). We tested differential 5mC and 5hmC associated with AD at these probes, controlling for age at death and cell-type proportion estimates. None of the differences in 5mC (minimum *P* = 2.47E-03) or 5hmC (minimum *P* = 1.53E-03) were significant when correcting for multiple testing (n = 6,838 tests; *P* < 7.31E-05), indicating that there is little direct overlap in AD-associated variation in H3K27ac and DNA modifications. Furthermore, comparing effect sizes at these 6,838 peak–probe pairs identified no evidence for a correlation between AD-associated H3K27ac and 5mC differences (r = 0.009, *P* = 0.443; **Figure 5-20**) with a small, but significant, negative correlation for 5hmC (r = -0.045, *P* = 1.63E-04; **Figure 5-20**). As expected, both DNA modifications are significantly lower in the vicinity of H3K27ac peaks compared to the genome-wide 450K array background (5mC: *P* < 1.00E-50**,** beta difference = 12%, **Figure 5-20**; 5hmC: *P* = 3.61E-30, beta difference = 0.16%; **Figure 5-20**), consistent with H3K27ac being localized at active enhancers and promoters. Interestingly, peaks characterized by significantly hyperacetylated H3K27ac in AD are enriched for intermediate levels of 5mC (beta in [0.4,0.6]) compared to the array-wide background (*P* = 2.95E-07, hypergeometric test). Furthermore, peaks characterized by AD-associated hyperacetylation have increased 5hmC compared to hypoacetylated peaks (*P* < 1.00E-50, beta difference = 2.17%), the non-significant peaks (*P* = 3.12E-25, beta difference = 1.33%) and the whole array background (*P* = 5.04E-20, beta difference = 1.17%; **Figure 5-20**).

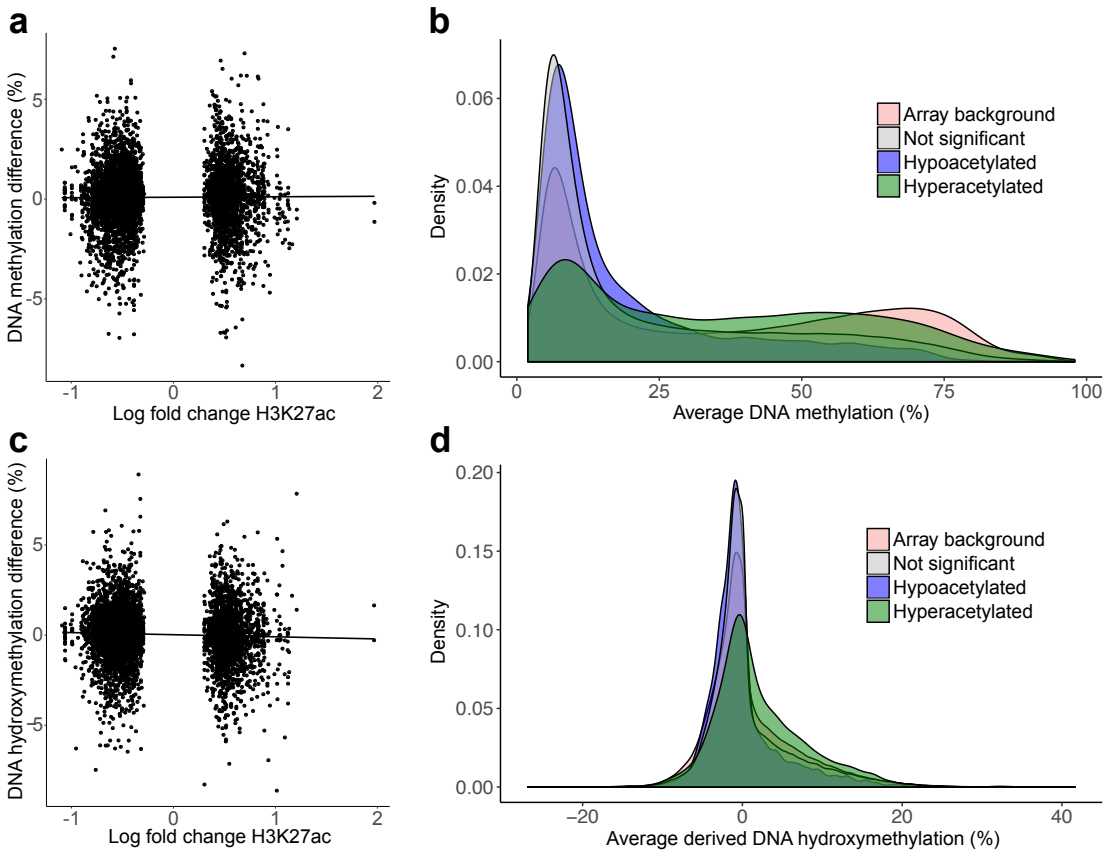


Figure 5-20. Diverging DNA methylation patterns are observed around AD hyperacetylated peaks. (**a**) Effect sizes for association between AD and DNA methylation at CpG probes within 1kb of FDR significant differentially acetylated peaks are not correlated with log fold change in H3K27ac at these peaks (*P* = 0.443). (**b**) Probes in vicinity of entorhinal cortex H3K27ac peaks show a skew towards lower DNA methylation compared to the array-background (*P* < 1.00E-50, beta difference = 12%), whereas hyperacetylated peaks show an enrichment in intermediate DNA methylation values, compared to the whole array background (*P* = 2.95E-07). (**c**) Effect sizes for association between AD and DNA hydroxymethylation at CpG probes within 1kb of FDR significant peaks are minimally but significantly negatively correlated with log fold change in H3K27ac at these peaks (r = -0.045, *P* = 1.63E-04). (**d**) DNA hydroxymethylation was derived by subtracting the true methylation percentage measured using oxidised bisulfite treatment from the combined methylation value assessed using regular bisulfite treatment. DNA hydroxymethlation is slightly lower near H3K27ac peaks, compared to the array background (*P* = 3.61E-30, beta difference = 0.16%), while probes in vicinity of AD hyperacetylated H3K27ac peaks show higher levels of DNA hydroxymethylation compared to those near hypoacetylated (*P* < 1.00E-50, beta difference = 2.17%), or not significantly differentially acetylated peaks (*P* = 3.12E-25, beta difference = 1.33%) and the whole array background (*P* = 5.04E-20, beta difference = 1.17%).

5.4 Discussion

In this study, we quantified genome-wide levels of H3K27ac in post-mortem entorhinal cortex tissue samples, identifying widespread AD-associated acetylomic variation. Strikingly, differentially acetylated peaks were identified in the vicinity of genes implicated in both tau and amyloid neuropathology (*MAPT*, *APP*, *PSEN1*, *PSEN2*), as well as genomic regions containing variants associated with sporadic late-onset AD (*CR1*, *TOMM40*). Both *MAPT* and *PSEN2* are characterized by an extended hyperacetylated region upstream of the transcription start site mapping to brain-related enhancers. We show that genes annotated to AD-associated hyper- and hypoacetylated peaks are enriched for brain- and neuropathology-related functions. This is first study of variable H3K27ac yet undertaken for AD, and in addition to identifying molecular pathways associated with AD neuropathology, introduces a framework for genome-wide studies of this modification in complex disease.

Given its close relationship with transcriptional activation, for example via the mediation of transcription factor binding, the identification of AD-associated variation in H3K27ac highlights potential novel regulatory genomic pathways involved in disease etiology. We find widespread alterations in H3K27ac associated with AD, including in the vicinity of several genes directly involved in the progression of neuropathology. Two main mechanisms are speculated to trigger pathogenesis in AD: Aβ and tau pathology (Selkoe and Hardy, 2016, Ittner and Gotz, 2011). Our results show disrupted regulatory variation at genes involved in both of these pathways (*APP*, *PSEN1*, *PSEN2*, *MAPT*), supporting the notion that both processes are involved in the onset of neuropathology. Interestingly, although we profile brains affected by sporadic late-onset AD, we identify widespread altered H3K27ac in the vicinity of genes implicated in familial early-onset AD. This indicates that these two forms of the disease may share common pathogenic pathways and mechanisms. Given that histone-acetylation modifiers are amongst the most promising target pharmacological treatments of AD (Graff and Tsai, 2013a, Fischer, 2014), the identification of altered H3K27ac in AD is important, giving clues as to which genes and pathways may be involved. Non-coding genetic variation, in form of acetylation-QTLs, may play a role in this context, in light of reports showing that variable H3K27ac regions are associated with genetic variation (Kasowski et al., 2013).

Our study has a number of limitations, which need to be considered when interpreting these results. First, we undertook ChIP-seq using “bulk” entorhinal cortex samples comprising a mix of neuronal and non-neuronal cell-types. This is an important limitation in studies of a disease characterized by cortical neuronal loss. However, we were able to control for variation in neuronal proportions in our samples by deriving neuronal proportion estimates for each sample using DNA methylation data generated on the same tissue samples using an established method (Guintivano et al., 2013). Second, our cross-sectional analysis of post-mortem brain tissue makes direct causal inference difficult, and it is plausible that the differences in H3K27ac we observe result from the AD pathology itself. In this regard, however, it is interesting that we see disease-associated H3K27ac in the vicinity of genes causally implicated in familial forms of AD. Third, we have assessed a relatively small number of samples. In this light, it is notable that we identify robust differences between AD cases and controls, with disease-associated regulatory variation in genes and functional pathways known to play a role in the onset and progression of neuropathology. The clear clustering between patients and controls at our differentially acetylated peaks suggests that despite a complex and heterogeneous etiology, AD is characterized by a common molecular pathology in the entorhinal cortex. Fourth, chromatin architecture and transcriptional regulation is influenced by a multitude of epigenetic mechanisms. Although H3K27ac profiles provide relatively robust information about transcriptional activity, they represent only one of perhaps ~100 post-translational modifications occurring at > 60 histone amino-acid residues regulating genomic function.

In summary, we provide compelling evidence for widespread acetylomic dysregulation in the entorhinal cortex in AD. Our data suggest that H3K27ac at multiple loci, including in the vicinity of several known AD risk genes – *APP*, *CR1*, *MAPT*, *PSEN1*, *PSEN2* and *TOMM40* – is robustly associated with disease, supporting the notion of common molecular pathways in both familial and sporadic AD. In addition to identifying molecular pathways associated with AD neuropathology, we present a framework for genome-wide studies of histone modifications in complex disease, integrating our data with results obtained from genome-wide association studies as well as other epigenetic marks profiled on the same samples.

6. General Discussion

6.1 Key findings

6.1.1 Severe psychosocial deprivation in early childhood is associated with increased DNA methylation across a region spanning the transcription start site of *CYP2E1*

The primary aim of my first empirical data chapter was to study methylomic correlates of severe early-life institutional deprivation. Using the English and Romanian Adoption Study (ERA), a unique longitudinal cohort established following a "natural experiment", I investigated the association between genome-wide patterns of DNA methylation quantified in buccal swab DNA of 49 individuals and the duration of deprivation exposure. A differentially methylated region (DMR) spanning 9 sequential CpG sites and overlapping the transcription start site of the cytochrome P450 gene *CYP2E1* was significantly hypermethylated in individuals exposed to extended periods of severe institutional deprivation (> 6 months). In addition to the deprivation related methylomic variation, I found a significant association between the same *CYP2E1* DMR and measures of social cognitive performance. These findings provide support for the hypothesis that severe early-life adverse exposures induce enduring epigenetic variation and potentially mediate the socio-cognitive sequelae of these stressful events.

6.1.2 There is no systematic association between early-life victimisation and epigenetic variation in whole blood

For my second empirical chapter, I set out to follow-up the association between severe early-life adversity and epigenetic variation by profiling DNA methylation in a large population-based cohort. To my knowledge this represents the most comprehensive analysis undertaken characterising epigenetic correlates of victimisation across the first two decades of life. Using whole blood DNA methylation profiles from 1,658 participants of the Environmental Risk (E-Risk) Longitudinal Twin Study, I interrogated methylomic variation associated with a variety of victimisation exposures in childhood (age 5-12) and adolescence (age 12-18). Overall, this study found little evidence for epigenetic variation in blood associated with victimisation exposure in childhood, adolescence or cumulatively. This study highlights both the importance of undertaking epidemiological studies in the disease-affected tissue and the importance of adequately controlling for potential confounders such as smoking.

6.1.3 Tissue-specific allelically-skewed DNA methylation is widespread and often associated with DNA sequence variation

The third empirical chapter of my PhD thesis focussed on profiling intra- and inter-individual patterns of allele-specific DNA methylation (ASM). Using matched post-mortem samples dissected from seven different brain regions and whole blood collected prior to death from multiple individuals, I characterised covariation of ASM between blood, cerebellum and cortex. This study uncovered the widespread and frequently tissue-specific occurrence of ASM across the human genome. In addition to confirming ASM at known imprinted regions, I identified examples of putative genotype-driven ASM, some of which are tissue-specific. This chapter highlighted how inter-individually variable ASM tends to be flanked by larger regions of intermediate DNA methylation and is enriched in DNase I hypersensitivity sites.

6.1.4 Widespread cortical dysregulation of H3K27ac is identified in Alzheimer's Disease, including at *MAPT* and *PSEN2*

In my final empirical chapter I characterised H3K27ac profiles in 47 entorhinal cortex samples from neuropathologically well-characterised individuals. The main aim of this study was to identify variation in histone-acetylation associated with Alzheimer's disease (AD). Using ChIP-seq, I profiled H3K27ac in post-mortem brain samples from patients affected by AD as well as neuropathology-free control samples. This chapter provides evidence for widespread disease-associated variation in H3K27ac, including at several genes previously implicated in tau and amyloid pathology (*APP*, *MAPT, PSEN1, PSEN2*). Using an integrated genetic and epigenetic approach I identified regions exhibiting converging evidence for disease-associated variation from genetic and epigenetic profiling. This is one of the first genome-wide studies of histone acetylation of any complex disease and the largest to date of AD and therefore also outlines a framework for genome-wide studies of histone modifications in disease.

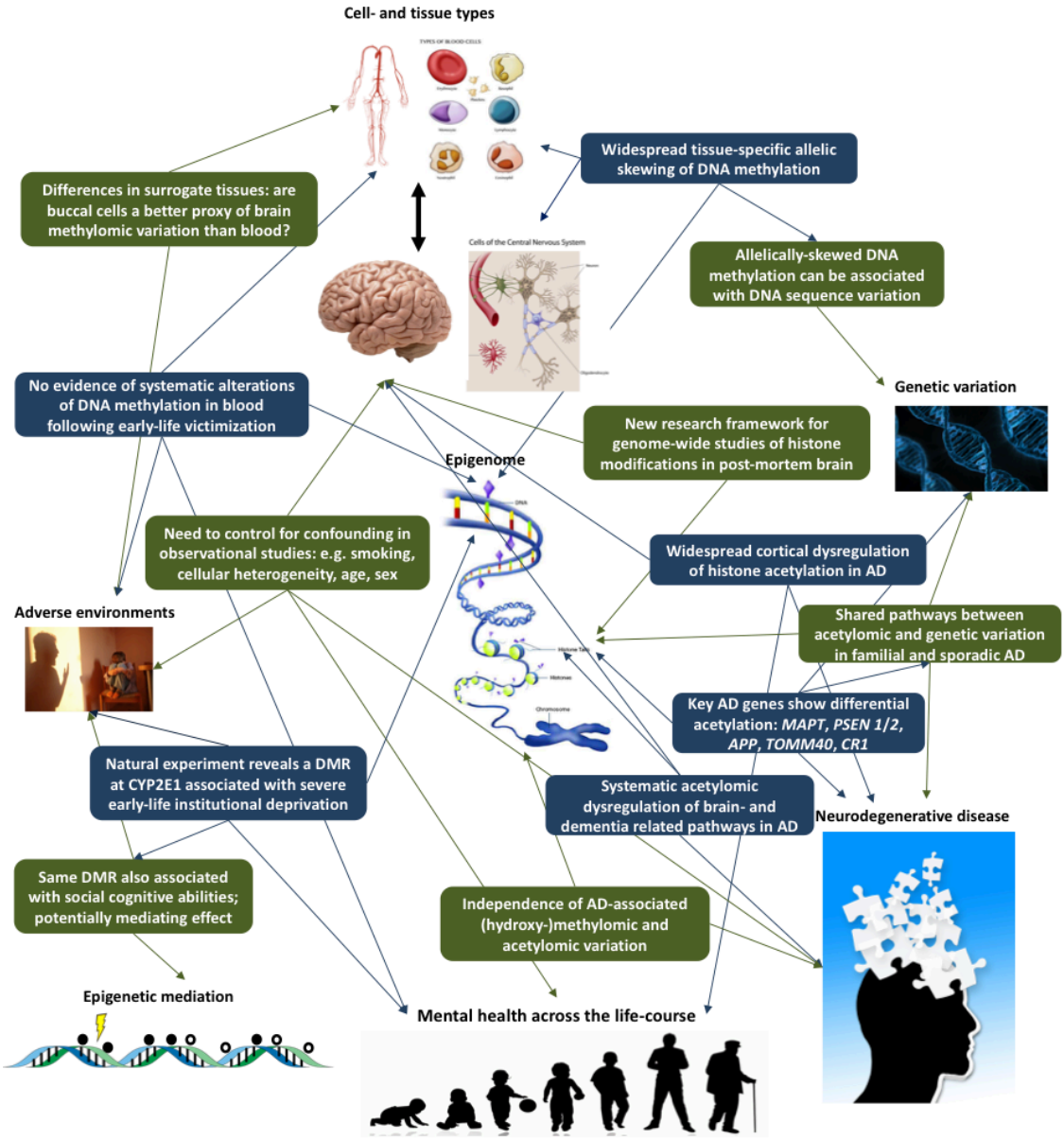


Figure 6-1. Summary of the key findings and implications of my PhD thesis, in relation to the research themes described in the introduction. Findings are coloured in blue, implications in green. For image credits see **Figure 1-10**.

6.2 Implications of my findings

During my PhD I employed a variety study designs, experimental approaches and technologies as well as analysis strategies to explore aspects of epigenetic variation in brain-related phenotypes. This thesis contributes to the field of epigenetic epidemiology of mental health in a number of important ways: First, two unique and complementary approaches were used to characterise epigenetic correlates of early-life adversity in surrogate tissue. Used together, these two approaches were able to overcome several challenges faced by previously published work in this field, including sample size concerns, confounding and

selection biases. Second, the thesis contributes to a growing body of work identifying brain-specific epigenetic dysregulation associated with Alzheimer's disease. Third, it presents one of the first genome-wide studies of H3K27ac, providing a framework for future epidemiological analyses involving histone modifications. Fourth, the work presented in this thesis enhances our understanding of tissue-specific epigenetic variation and the utility of surrogate tissues in epigenetic studies of psychiatric phenotypes. Fifth, this thesis presents novel insights into the interplay of genetic and epigenetic variation, both in the pathology-free brain and in the context of AD-associated dysregulation. Key findings and implications of my PhD thesis are summarised in **Figure 6-1**.

6.2.1 Epigenetic correlates of adverse environments

Several previously published studies have reported variation in DNA methylation regulating the function of genes of the hypothalamic-pituitary-adrenal (HPA) axis in consequence to early-life stress and adversity. Using both animal and human samples, these studies have focused primarily on candidate genes, including *NR3C1*, *FKBP5* and *SLC6A4*. **Chapters 2** and **3** of my thesis used two complementary designs, a unique natural experiment and a large-population based cohort including twins to further investigate the association between early-life adversity and DNA methylation in peripheral tissue. Using Illumina 450K Human Methylation microarrays and bioinformatics approaches, DNA methylation was quantified at ~ 450,000 CpG sites in the human genome in buccal cells (**Chapter 2**) or whole blood (**Chapter 3**). Together, these chapters allowed me to address the question of epigenetic signatures of early-life adversity in two important ways: Using a natural experiment focussing on severe early-life stress, **Chapter 2** interrogates methylation associated with the extreme adversity experienced in the orphanages of Ceausescu's Romania. **Chapter 3** provides comprehensive insights from the largest study to date of early-life victimisation, using a UK-representative population-based sample and detailed longitudinal phenotypic data recorded across the first two decades of life. In addition, **Chapter 3** exploits the twin design of the E-Risk Longitudinal Study, allowing me to control for confounding effects caused by DNA sequence variation and/or shared family environments. While the two studies have diverging results, neither of them identified DNA methylation variation in any of previously reported candidate

genes. However, a striking association was identified in a DMR overlapping the promoter and TSS of *CYP2E1* in **Chapter 2**. Methylation at this DMR was additionally associated with measures of social cognition. In **Chapter 3** I did not observe significant associations between DNA methylation and early-life adversity beyond those confounded by smoking. The non-replication of the *CYP2E1* finding could be related to a number of differences between the studies: First, the tissues used were different, and there is a possibility that DNA methylation in whole blood is not an ideal surrogate for neuroepigenetic variation. Second, the exposure experienced in ERA is more extreme and less heterogeneous – i.e. severe global institutional deprivation - and might not be directly comparable with the victimisation exposures characterised in E-Risk. Third the exposure timing in ERA is focused on the first four years of life, while victimisation profiling in E-Risk begins at age 5. There is a possibility that the first few years of life represent a developmentally more sensitive period and that adverse experiences in this period has a more dramatic impact on long-term biological outcomes.

6.2.2 AD associated epigenetic variation

Previous work in epigenetic epidemiology of brain disorders has focussed almost exclusively on DNA methylation, often profiled in peripheral tissues. However, the human epigenome is highly complex and tissue-specific (Kundaje et al., 2015, Marzi et al., 2016) and surrogate tissues may not be adequate proxies for molecular biological processes occurring in and possibly causally contributing to neuropsychiatric disease. There is a growing body of research implicating brain-specific dysregulation of DNA methylation in Alzheimer's disease neuropathology (Lunnon et al., 2014, De Jager et al., 2014). The work presented in **Chapter 5** of this PhD thesis contributes to this existing body of research by characterising AD-associated variation in the histone modification H3K27ac in post-mortem entorhinal cortex in a highly novel approach: a genome-wide association study of histone acetylation. I found widespread patterns of acetylomic dysregulation associated with AD in this study, enriched in several disease- and neurotransmitter-specific biological pathways. Unlike results from previously published work based on DNA methylation data, two of the most striking findings from this study implicate hyperacetylation clusters associated with *MAPT* and

PSEN2, two well-established AD-related genes in the disease. Intriguingly, both clusters are characterised by brain-related active chromatin states. Furthermore, acetylomic dysregulation is observed around other well-known AD-associated genes including *APP* and *PSEN1* (familial, early-onset AD), as well as *TOMM40* and *CR1* (sporadic AD; (Lambert et al., 2013)). My work also shows that regions of AD-associated differential acetylation and DNA modifications (5mC and 5hmC) are largely non-overlapping, and therefore this H3K27ac-EWAS provides unique novel insights into epigenetic dysregulation in Alzheimer' disease. Given, the clear enrichment of brain-specific regulatory pathways and functions of my results, these findings have the potential to advance our understanding of the pathways by which epigenetic drugs function: HDAC inhibitors are known to be potent candidate drugs in neurodegenerative disease and show pervasive effects on synaptic plasticity and memory formation (Guan et al., 2009, Fischer et al., 2007, Kilgore et al., 2010), possibly affecting specific AD-dysregulated histone-acetylation clusters.

6.2.3 Genome-wide epidemiological studies of histone modifications

In addition to strengthening the evidence for a role of epigenetic dysregulation in AD, **Chapter 5** also represents one of the first large-scale genome-wide studies of histone acetylation in a complex disease. While histone-modifications have been profiled across numerous human cell- and tissue-types in recent years (Kundaje et al., 2015), most of the work has focused on characterising histone modification patterns from individual pathology-free samples. A first genome-wide study of H3K27ac was published for autism spectrum disorder last year (Sun et al., 2016). Reassuringly, the cortical H3K27ac peaks identified by Sun and colleagues - as well as brain-specific H3K27ac peaks from the NIH Roadmap Epigenomic Consortium - overlap those identified in entorhinal cortex samples profiled in **Chapter 5**. Building on the work from Sun and colleagues, my analyses demonstrate the suitability of post-mortem brain tissue for histone modification profiling using ChIP-seq. A research framework for this type of study is described, including stringent laboratory protocols, automated immunoprecipitation, data pre-processing and quality control (**Figure 6-2**). I also addressed the appropriate statistical modelling of ChIP-seq data, including controls for potential biological (age, cell-type composition) and technical covariates (library size, dispersion).

Future epidemiological studies of histone modifications will hopefully benefit from the experimental and analytical framework described in my analysis.

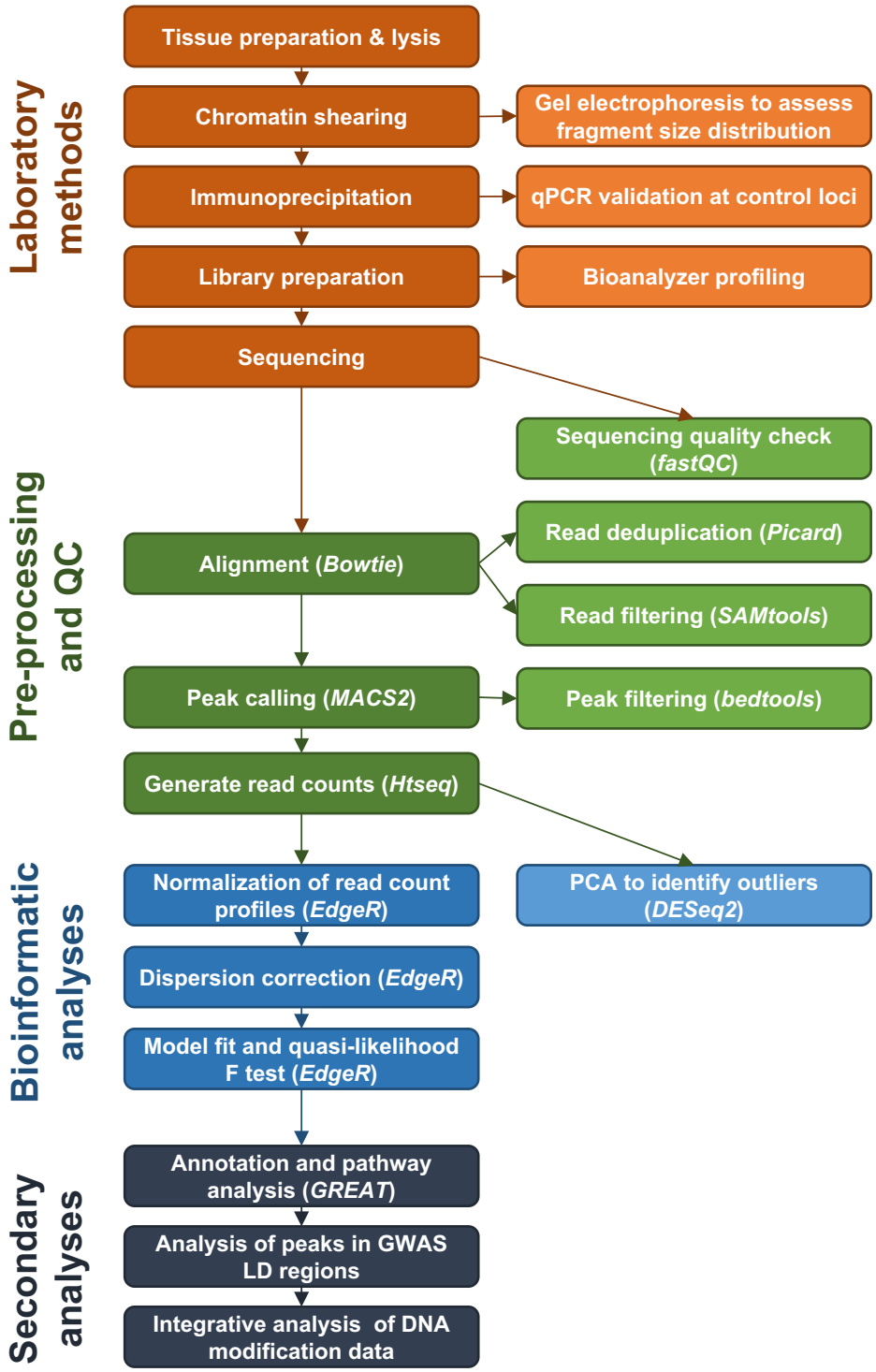


Figure 6-2. Overview of the research framework developed for genome-wide studies of histone modifications.

6.2.4 Tissue-specificity of epigenetic modifications

Chapter 4 characterised tissue-specific profiles of allelically-skewed DNA methylation across human brain and blood, adding to published tissue-specific maps of epigenetic modifications (Kundaje et al., 2015, Hannon et al., 2015), but importantly also considering intra-individual epigenetic covariation across tissues (Davies et al., 2012, Hannon et al., 2015). In addition to confirming allelically-skewed DNA methylation in the vicinity of known imprinted regions, my results highlight putative genetically-driven instances of allelic-skewing, some of which are highly tissue-specific. The comprehensive profiling of tissue-specific genetic effects on epigenetic states is of particular importance in uncovering the molecular pathways that connect genetic variation with phenotypic outcomes in disease-affected tissues (Meaburn et al., 2010) and the work presented in **Chapter 4** contribute an important piece to this puzzle.

Chapters 2, 3 and **5** profile exposure- and disease-associated epigenetic variation collected from three different sources: buccal swabs, whole blood and post-mortem brain. Previous work in AD has highlighted tissue-specific dysregulation of DNA methylation, with disease-associated differentially methylated probes (DMPs) identified only in the cortex but not in the pathology-free cerebellum or whole blood of AD patients (Lunnon et al., 2014). In line with these findings, we observe widespread AD-associated variation in H3K27ac in the disease-affected entorhinal cortex. Furthermore, some of our key findings are characterised by cortex-specific active chromatin states and histone-acetylation profiles, suggesting that these regulatory marks may also be cortex-specific. Results for epigenetic correlates of adverse psychosocial environments in surrogate tissues are more mixed: In **Chapter 2**, I was able to identify regional differences in DNA methylation associated with severe institutional deprivation in buccal cells. On the other hand, I found no robust associations between early-life victimisation and DNA methylation in whole blood in **Chapter 3**. This may indicate that buccal cells could be a better proxy of brain epigenomic variation than blood. Nonetheless, epigenetic profiling of blood in epidemiological contexts can be a valuable biomarker of other factors, as illustrated by the pervasive associations identified with cigarette smoking.

6.2.5 Genetic and epigenetic variation

For my PhD thesis, I applied a variety of techniques (both experimental and analytical) to study multiple epigenetic phenomena across a range of health questions related to the brain. Importantly, epigenetic mechanisms do not act independently of the genetic sequence or other epigenetic modifications, even though they are generally interrogated individually. Throughout my thesis, I have employed a variety of strategies that aimed to address (epi-)genetic variation in an integrated manner. First, **Chapter 4** addresses the question of allelic imbalances in DNA methylation across the human genome, most of which are genetically driven. To characterise this phenomenon, epigenetic and genetic laboratory methods were combined: a cocktail of methylation-sensitive restriction enzymes was used to digest DNA prior to hybridisation to a genotyping microarray. Relative genotype signal intensity changes were used as a measure to quantify allelic-skew in DNA methylation, exploiting heterozygosity to discern between methylation on the two alleles. The resulting tissue-specific ASM maps have important implications for genetic studies of psychiatric disease as mediating or interacting factors (Meaburn et al., 2010). Second, I exploit the fact that monozygotic twins are genetically identical, to explore methylomic correlates of victimisation exposure while controlling for genetic effects in **Chapter 3**. Twin studies offer one of the most powerful designs in epigenetic epidemiology controlling for several confounding factors including familial environment and genetic effects (Castillo-Fernandez et al., 2014, Mill and Heijmans, 2013). In the E-Risk cohort, methylomic effects at smoking-confounded victimisation associated probes are attenuated in a twin-difference model, even when not controlling directly for smoking-differences between twins, despite reported differences in smoking behaviour. Finally, **Chapter 5** uses a multitude of public and additionally collected data sources to explore genetic and epigenetic variation implicated in AD in an integrated approach. To supplement the primary analysis focussing on disease-associated H3K27ac variation, I examined linkage disequilibrium blocks, to explore epigenetic variation in genomic regions associated with disease, characterising acetylomic variation in these regions. Using DNA methylation and hydroxymethylation profiles from the same samples, I find that regions exhibiting AD-associated differential acetylation and DNA modifications are largely non-overlapping, but that disease-associated hyper-

and hypoacetylated regions are characterised by different profiles of 5mC and 5hmC.

6.3 Limitations and future research directions

As outlined in the introduction, a perfect study design does not exist in epigenetic epidemiology of brain disorders, and the studies presented in this thesis are subject to a number of methodological limitations. These limitations are discussed in detail at the end of each chapter, but several recurring limitations affect the data presented in many of my chapters and will be discussed in the broader context of this thesis discussion.

6.3.1 Sample size and replication

Some of the studies presented in this PhD thesis are somewhat limited in sample size, usually for specific reasons: The work conducted for **Chapter 2** focuses on a unique small sample, a subset of the English and Romanian Adoption Study. Replication of the results is warranted, but similar samples exposed to severe adversity are scarce and clearly cannot be repeated experimentally for ethical reasons. Follow-up including the full cohort of ERA is planned and could be complemented by replication in larger population-based studies. Although we profiled a wide range of brain regions and blood from the same donors, the number of samples used in **Chapter 4** is very limited, making definitive statements about genetically-driven methylomic variation or general profiles of inter-individual variation difficult. Larger studies of ASM, in particular in combination with profiling of methylation and expression QTLs, would provide invaluable new maps and insights about genetic, methylomic and transcriptional covariation. The work presented in **Chapter 5** constitutes the largest study of histone acetylomic variation in AD and in the entorhinal cortex to date. Post-mortem brain samples are not as readily available as peripheral tissue, but brain banking is increasingly enabling research into the molecular biology of the human brain (Kretzschmar, 2009); meanwhile the costs associated with ChIP-seq, if employed in a larger epidemiological study design, are still substantial. However, the widespread acetylomic differences observed in my study and, and particularly

those in several known AD genes, are extremely promising and encourage future research on larger cohorts.

6.3.2 Tissue and cell-types

The two studies interrogating early-life adversity (**Chapters 2** and **3**) were conducted on peripheral tissue samples. With very few exceptions (e.g., suicide completer brain samples), studies of these types of exposures will always have to be based on surrogate tissue, in particular if part of larger longitudinal cohorts. Given my results and previous reports, it is possible that buccal cells may be a more appropriate surrogate for brain methylomic variation (Lowe et al., 2013), but the profiled ERA sample was limited in size and the presented results require replication. Additional and more accurate maps of intra-individual epigenetic variation across tissues are required, but buccal cells are not yet generally collected in the context of brain banking. Work detailing blood-brain methylomic covariation across a larger sample has been published previously (Hannon et al., 2015) and my work outlined in **Chapter 4** contributes to the collection of blood-brain epigenomic maps (Marzi et al., 2016). Despite the lack of significant associations between victimisation and blood methylomic variation in **Chapter 3**, whole blood has immense potential as an epidemiological biomarker for other exposures, including smoking behaviour (Joehanes et al., 2016).

The second big issue with regards to cell- and tissue-types is cellular heterogeneity, and its potential to confound epigenetic association studies. Reference-based estimators for blood (used in **Chapter 3**) and brain (used in **Chapter 5**) cell-type compositions have been published and shown to perform well (Houseman et al., 2012, Guintivano et al., 2013). However, the CETS algorithm for brain cell-type estimates only distinguishes between neurons and non-neuronal cells, while more detailed subtyping of neurons, or identification of specific non-neuronal cell-types like microglia may be of interest in the context of some psychiatric diseases. Furthermore, the reference dataset, on which the CETS algorithm was fitted, was based on NeuN immunolabelling, which fails to capture some types of neurons, including cerebellar Purkinje cells and Golgi cells (Jeffries and Mill, 2017). In addition to estimation error for both reference-based estimators, these post-hoc statistical approaches also may not be able to pick up epigenetic differences in single cell-subtypes, due to a convolution of signal,

meaning that a significant epigenetic difference in a single cell-type may be drowned out by the epigenetic noise produced by all other cell-types occurring in the heterogeneous tissue sample. Single cell-type extraction methods, including fluorescent activated cell sorting (FACS) or laser capture microdissection (LCM), would be able to target questions of single cell-type epigenomic variation, but are highly work-intense and not feasible on previously isolated DNA samples.

6.3.3 Causality and mechanism

All the analyses presented in this thesis (**Chapters 2**, **3** and **5**) are cross-sectional, studying epigenetic variation across multiple individuals collected at a single time. This makes attribution of causal and even temporal relationships between exposures, phenotypic and epigenetic variation elusive. In **Chapter 2** we use a natural experiment, allowing us to rule out effects of genetics or external factors on the institutional exposure duration. Assuming, that institutional deprivation lead to the observed epigenetic differences, we still cannot conclude that the psychosocial components of the exposure caused the variation in DNA methylation. While we ascertained that smoking and medication use were unrelated to the exposure duration, other factors pertaining to the institutional environment, such as environmental toxins, could have contributed to the association with DNA methylation. In contrast, in **Chapter 3** we find that the association between victimisation exposure and DNA methylation is confounded by smoking. One possible explanation is that socioeconomic status increases both the likelihood of victimisation exposure and smoking. Planned subsequent repeated profiling of DNA methylation in these two longitudinal cohorts will give us better insights into the temporal profiles of epigenetic variation and effects of new environmental exposures.

The widespread acetylomic associations with AD reported in **Chapter 5** are very encouraging, implicating known AD genes and highly disease- and brain-specific pathways, further establishing them as candidates for a mechanistic role in AD aetiology and progression. However, given the cross-sectional setup of this study, it is completely unknown whether the epigenetic dysregulation predates the disease or arises as a consequence. Epigenetic variation could causally contribute to AD, but could also reflect a downstream effect of the disease or medication, or a by-product relating to other factors. Future functional work will

be required to shed more light onto these associations. For example, given the dysregulated acetylation cluster upstream of *MAPT* is covered by enhancers, methods profiling 3D chromatin structure, such as chromatin conformation capture (3C) or enhancer luciferase assays could be used to profile the genes it interacts with. Furthermore, CRISPR-mediated knockouts in mouse models or cell-culture could be employed to see how disrupting the enhancer affects transcription of *MAPT*. In addition to functional experimental approaches, genotypic profiling would open up the possibility of Mendelian randomisation for statistical causal inference. However, to optimise power for the detection of mQTLs, larger samples than those presented in **Chapter 2** and **5** would be required.

6.3.4 Genetics and data integration

While indirect approaches involving genetic variation have been described in **Chapters 3**, **4** and **5**, direct profiling of genotypic variation in the context of epigenetic association studies would be useful, enabling a more detailed analysis of genetic and epigenetic contributions to the epidemiological measures, and efforts to follow-up the work presented in **Chapters 2**, **3** and **5** with genotypic information are indeed ongoing. One potential benefit resulting from genotyping is the interrogation of acetylomic variation not only in regions of risk variants but in their actual context. For example, do carriers of the *APOE*-ε4 risk variant for Alzheimer's disease exhibit different profiles of histone acetylation in the surrounding genomic region? A second line of investigation could look into allelic variation in H3K27ac. Using genotypic information of the samples characterised in **Chapter 5**, a haplotyped genome could be assembled, to which ChIP-seq reads would be aligned allele-specifically. This would enable us to assess naturally occurring allelic biases in histone acetylation across the entorhinal cortex (using only the pathology-free samples), as well as the identification of disease-specific allelic-biases in H3K27ac (using all samples). Genotyping in the longitudinal cohorts examined in **Chapters 2** and **3** will hopefully allow us to tease apart genetic and environmentally induced epigenetic changes in greater detail. Using both risk variants but importantly also polygenic risk scores (PRS) for mental health outcomes observed following adversity, genetic contributions to these outcomes could be better accounted for (Hannon et al., 2016a). Another

important methodological advantage of collecting genotypic information on the participants of these studies would be to rule out effects of population-stratification, where population-specific genotypic variation correlates both with the exposure or outcome measures as well as with the associated epigenetic variation.

6.3.5 Technology

Most technologies used in epigenetic epidemiology have certain limitations, including technical noise, genomic coverage/resolution or sample requirements. The Illumina 450K Human Methylation array used to profile DNA methylation following adversity in **Chapters 2** and **3** is limited to profiling methylation at ~450,000 CpG sites, a fraction of the total 28 million sites found across the human genome. While it is purposefully designed to cover CpG sites annotated to most protein-coding genes it lacks broader coverage of identified regulatory elements. Additionally, traditional sodium-bisulfite conversion, as employed in **Chapters 2** and **3**, is unable to distinguish between the two cytosine modifications 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC). While this should not present a serious problem for these two studies, given the extremely low prevalence of 5hmC in most non-neuronal tissues, it needs to be considered when profiling DNA modifications across the brain: the additional DNA methylation and hydroxymethylation data presented in **Chapter 5**, used oxidative bisulfite-conversion, making the distinction between true 5mC and 5hmC possible.

6.4 Final words

Psychiatric epigenetics is a rapidly advancing field with many encouraging developments and findings in recent years. Researchers are beginning to link regulatory variation with genetic and environmental risk factors associated with neuropsychiatric phenotypes. My PhD explored epigenetic variation in the context of psychosocial environments and neuropsychiatric health and disease employing novel technologies and unique study designs to shed light onto how genetic and environmental factors influence risk of psychiatric disease. My findings encourage further work, inferring causal, functional and mechanistic

relationships of epigenetic signatures of mental health and neuropathology, which will hopefully eventually lead to novel therapeutic and pharmacological interventions, ameliorating adverse mental health sequelae following early-life stress and preventing or reversing the processes leading to neurodegeneration.

References

ABERG, K. A., MCCLAY, J. L., NERELLA, S., CLARK, S., KUMAR, G., CHEN, W., KHACHANE, A. N., XIE, L., HUDSON, A., GAO, G., HARADA, A., HULTMAN, C. M., SULLIVAN, P. F., MAGNUSSON, P. K. & VAN DEN OORD, E. J. 2014. Methylome-wide association study of schizophrenia: identifying blood biomarker signatures of environmental insults. *JAMA Psychiatry,* 71**,** 255-64.

ADEGBOLA, A. A., COX, G. F., BRADSHAW, E. M., HAFLER, D. A., GIMELBRANT, A. & CHESS, A. 2015. Monoallelic expression of the human FOXP2 speech gene. *Proc Natl Acad Sci U S A,* 112**,** 6848-54.

AKBARIAN, S. & HUANG, H. S. 2009. Epigenetic regulation in human brain-focus on histone lysine methylation. *Biol Psychiatry,* 65**,** 198-203.

AKBARIAN, S., RUEHL, M. G., BLIVEN, E., LUIZ, L. A., PERANELLI, A. C., BAKER, S. P., ROBERTS, R. C., BUNNEY, W. E., JR., CONLEY, R. C., JONES, E. G., TAMMINGA, C. A. & GUO, Y. 2005. Chromatin alterations associated with down-regulated metabolic gene expression in the prefrontal cortex of subjects with schizophrenia. *Arch Gen Psychiatry,* 62**,** 829-40.

ALEXANDER, N., WANKERL, M., HENNIG, J., MILLER, R., ZANKERT, S., STEUDTE-SCHMIEDGEN, S., STALDER, T. & KIRSCHBAUM, C. 2014. DNA methylation profiles within the serotonin transporter gene moderate the association of 5-HTTLPR and cortisol stress reactivity. *Transl Psychiatry,* 4**,** e443.

AMIR, R. E., VAN DEN VEYVER, I. B., WAN, M., TRAN, C. Q., FRANCKE, U. & ZOGHBI, H. Y. 1999. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Genet,* 23**,** 185-8.

ANDA, R. F., FELITTI, V. J., BREMNER, J. D., WALKER, J. D., WHITFIELD, C., PERRY, B. D., DUBE, S. R. & GILES, W. H. 2006. The enduring effects of abuse and related adverse experiences in childhood. A convergence of evidence from neurobiology and epidemiology. *Eur Arch Psychiatry Clin Neurosci,* 256**,** 174-86.

ANDERS, S., PYL, P. T. & HUBER, W. 2015. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics,* 31**,** 166-9.

ANDERSEN, S. L., TOMADA, A., VINCOW, E. S., VALENTE, E., POLCARI, A. & TEICHER, M. H. 2008. Preliminary evidence for sensitive periods in the effect of childhood sexual abuse on regional brain development. *J Neuropsychiatry Clin Neurosci,* 20**,** 292-301.

ANDREWS, S. 2010. FastQC: a quality control tool for high throughput sequence data.

ARSENEAULT, L., WALSH, E., TRZESNIEWSKI, K., NEWCOMBE, R., CASPI, A. & MOFFITT, T. E. 2006. Bullying victimization uniquely contributes to adjustment problems in young children: a nationally representative cohort study. *Pediatrics,* 118**,** 130-8.

ASHBURNER, M., BALL, C. A., BLAKE, J. A., BOTSTEIN, D., BUTLER, H., CHERRY, J. M., DAVIS, A. P., DOLINSKI, K., DWIGHT, S. S., EPPIG, J. T., HARRIS, M. A., HILL, D. P., ISSEL-TARVER, L., KASARSKIS, A., LEWIS, S., MATESE, J. C., RICHARDSON, J. E., RINGWALD, M., RUBIN, G. M. & SHERLOCK, G. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet,* 25**,** 25-9.

AVNER, P. & HEARD, E. 2001. X-chromosome inactivation: counting, choice and initiation. *Nat Rev Genet,* 2**,** 59-67.

BACCARELLI, A., WRIGHT, R. O., BOLLATI, V., TARANTINI, L., LITONJUA, A. A., SUH, H. H., ZANOBETTI, A., SPARROW, D., VOKONAS, P. S. & SCHWARTZ, J. 2009. Rapid

DNA methylation changes after exposure to traffic particles. *Am J Respir Crit Care Med,* 179**,** 572-8.

BALDWIN, J. R., ARSENEAULT, L., ODGERS, C., BELSKY, D. W., MATTHEWS, T., AMBLER, A., CASPI, A., MOFFITT, T. E. & DANESE, A. 2016. Childhood Bullying Victimization and Overweight in Young Adulthood: A Cohort Study. *Psychosom Med,* 78**,** 1094-1103.

BARTEL, D. P. 2009. MicroRNAs: target recognition and regulatory functions. *Cell,* 136**,** 215-33.

BATES, D., MAECHLER, M., BOLKER, B. & WALKER, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software,* 67**,** 1-48.

BEACH, S. R., BRODY, G. H., TODOROV, A. A., GUNTER, T. D. & PHILIBERT, R. A. 2010. Methylation at SLC6A4 is linked to family history of child abuse: an examination of the Iowa Adoptee sample. *Am J Med Genet B Neuropsychiatr Genet,* 153B**,** 710-3.

BEACH, S. R., BRODY, G. H., TODOROV, A. A., GUNTER, T. D. & PHILIBERT, R. A. 2011. Methylation at 5HTT mediates the impact of child sex abuse on women's antisocial behavior: an examination of the Iowa adoptee sample. *Psychosom Med,* 73**,** 83-7.

BELL, A. C. & FELSENFELD, G. 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature,* 405**,** 482-5.

BELL, J. T., PAI, A. A., PICKRELL, J. K., GAFFNEY, D. J., PIQUE-REGI, R., DEGNER, J. F., GILAD, Y. & PRITCHARD, J. K. 2011. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol,* 12**,** R10.

BERGER, S. L. 2007. The complex language of chromatin regulation during transcription. *Nature,* 447**,** 407-12.

BIRD, A. P. 1980. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res,* 8**,** 1499-504.

BIRNEY, E., SMITH, G. D. & GREALLY, J. M. 2016. Epigenome-wide Association Studies and the Interpretation of Disease -Omics. *PLoS Genet,* 12**,** e1006105.

BIRNEY, E., STAMATOYANNOPOULOS, J. A., DUTTA, A., GUIGO, R., GINGERAS, T. R., MARGULIES, E. H., WENG, Z., SNYDER, M., DERMITZAKIS, E. T., THURMAN, R. E., KUEHN, M. S., TAYLOR, C. M., NEPH, S., KOCH, C. M., ASTHANA, S., MALHOTRA, A., ADZHUBEI, I., GREENBAUM, J. A., ANDREWS, R. M., FLICEK, P., BOYLE, P. J., CAO, H., CARTER, N. P., CLELLAND, G. K., DAVIS, S., DAY, N., DHAMI, P., DILLON, S. C., DORSCHNER, M. O., FIEGLER, H., GIRESI, P. G., GOLDY, J., HAWRYLYCZ, M., HAYDOCK, A., HUMBERT, R., JAMES, K. D., JOHNSON, B. E., JOHNSON, E. M., FRUM, T. T., ROSENZWEIG, E. R., KARNANI, N., LEE, K., LEFEBVRE, G. C., NAVAS, P. A., NERI, F., PARKER, S. C., SABO, P. J., SANDSTROM, R., SHAFER, A., VETRIE, D., WEAVER, M., WILCOX, S., YU, M., COLLINS, F. S., DEKKER, J., LIEB, J. D., TULLIUS, T. D., CRAWFORD, G. E., SUNYAEV, S., NOBLE, W. S., DUNHAM, I., DENOEUD, F., REYMOND, A., KAPRANOV, P., ROZOWSKY, J., ZHENG, D., CASTELO, R., FRANKISH, A., HARROW, J., GHOSH, S., SANDELIN, A., HOFACKER, I. L., BAERTSCH, R., KEEFE, D., DIKE, S., CHENG, J., HIRSCH, H. A., SEKINGER, E. A., LAGARDE, J., ABRIL, J. F., SHAHAB, A., FLAMM, C., FRIED, C., HACKERMULLER, J., HERTEL, J., LINDEMEYER, M., MISSAL, K., TANZER, A., WASHIETL, S., KORBEL, J., EMANUELSSON, O., PEDERSEN, J. S., HOLROYD, N., TAYLOR, R., SWARBRECK, D., MATTHEWS, N., DICKSON, M. C., THOMAS, D. J., WEIRAUCH, M. T., GILBERT, J., et al. 2007. Identification and analysis of

functional elements in 1% of the human genome by the ENCODE pilot project. *Nature,* 447**,** 799-816.

BOKS, M. P., DE JONG, N. M., KAS, M. J., VINKERS, C. H., FERNANDES, C., KAHN, R. S., MILL, J. & OPHOFF, R. A. 2012. Current status and future prospects for epigenetic psychopharmacology. *Epigenetics,* 7**,** 20-8.

BOKS, M. P., VAN MIERLO, H. C., RUTTEN, B. P., RADSTAKE, T. R., DE WITTE, L., GEUZE, E., HORVATH, S., SCHALKWYK, L. C., VINKERS, C. H., BROEN, J. C. & VERMETTEN, E. 2015. Longitudinal changes of telomere length and epigenetic age related to traumatic stress and post-traumatic stress disorder. *Psychoneuroendocrinology,* 51**,** 506-12.

BOOIJ, L., SZYF, M., CARBALLEDO, A., FREY, E. M., MORRIS, D., DYMOV, S., VAISHEVA, F., LY, V., FAHEY, C., MEANEY, J., GILL, M. & FRODL, T. 2015. DNA methylation of the serotonin transporter gene in peripheral cells and stress-related changes in hippocampal volume: a study in depressed patients and healthy controls. *PLoS One,* 10**,** e0119061.

BOOTH, M. J., BRANCO, M. R., FICZ, G., OXLEY, D., KRUEGER, F., REIK, W. & BALASUBRAMANIAN, S. 2012. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science,* 336**,** 934-7.

BOOTH, M. J., OST, T. W., BERALDI, D., BELL, N. M., BRANCO, M. R., REIK, W. & BALASUBRAMANIAN, S. 2013. Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. *Nat Protoc,* 8**,** 1841-51.

BORGHOL, N., SUDERMAN, M., MCARDLE, W., RACINE, A., HALLETT, M., PEMBREY, M., HERTZMAN, C., POWER, C. & SZYF, M. 2012. Associations with early-life socio-economic position in adult DNA methylation. *Int J Epidemiol,* 41**,** 62-74.

BOYES, J. & BIRD, A. 1991. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell,* 64**,** 1123-34.

BRAAK, H. & BRAAK, E. 1991. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol,* 82**,** 239-59.

BRADLEY, R. G., BINDER, E. B., EPSTEIN, M. P., TANG, Y., NAIR, H. P., LIU, W., GILLESPIE, C. F., BERG, T., EVCES, M., NEWPORT, D. J., STOWE, Z. N., HEIM, C. M., NEMEROFF, C. B., SCHWARTZ, A., CUBELLS, J. F. & RESSLER, K. J. 2008. Influence of child abuse on adult depression: moderation by the corticotropin-releasing hormone receptor gene. *Arch Gen Psychiatry,* 65**,** 190-200.

BRESLOW, N. E. & DAY, N. E. 1980. Statistical methods in cancer research. Volume I - The analysis of case-control studies. *IARC Sci Publ***,** 5-338.

BRIND'AMOUR, J., LIU, S., HUDSON, M., CHEN, C., KARIMI, M. M. & LORINCZ, M. C. 2015. An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. *Nat Commun,* 6**,** 6033.

BROOKMEYER, R., JOHNSON, E., ZIEGLER-GRAHAM, K. & ARRIGHI, H. M. 2007. Forecasting the global burden of Alzheimer's disease. *Alzheimers Dement,* 3**,** 186-91.

BROWN, D. C., GRACE, E., SUMNER, A. T., EDMUNDS, A. T. & ELLIS, P. M. 1995. ICF syndrome (immunodeficiency, centromeric instability and facial anomalies): investigation of heterochromatin abnormalities and review of clinical outcome. *Hum Genet,* 96**,** 411-6.

BROWN, M. B. 1975. 400: A Method for Combining Non-Independent, One-Sided Tests of Significance. *Biometrics,* 31**,** 987–992.

BROWN, S. L., BIRCH, D. A. & KANCHERLA, V. 2005. Bullying perspectives: experiences, attitudes, and recommendations of 9- to 13-year-olds attending health education centers in the United States. *J Sch Health,* 75**,** 384-92.

BU, G. 2009. Apolipoprotein E and its receptors in Alzheimer's disease: pathways, pathogenesis and therapy. *Nat Rev Neurosci,* 10**,** 333-44.

BUCKLEY, N. J. 2007. Analysis of transcription, chromatin dynamics and epigenetic changes in neural genes. *Prog Neurobiol,* 83**,** 195-210.

BUITING, K., SAITOH, S., GROSS, S., DITTRICH, B., SCHWARTZ, S., NICHOLLS, R. D. & HORSTHEMKE, B. 1995. Inherited microdeletions in the Angelman and Prader-Willi syndromes define an imprinting centre on human chromosome 15. *Nat Genet,* 9**,** 395-400.

CASPI, A., SUGDEN, K., MOFFITT, T. E., TAYLOR, A., CRAIG, I. W., HARRINGTON, H., MCCLAY, J., MILL, J., MARTIN, J., BRAITHWAITE, A. & POULTON, R. 2003. Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science,* 301**,** 386-9.

CASSIDY, S. B. & SCHWARTZ, S. 1998. Prader-Willi and Angelman syndromes. Disorders of genomic imprinting. *Medicine (Baltimore),* 77**,** 140-51.

CASTILLO-FERNANDEZ, J. E., SPECTOR, T. D. & BELL, J. T. 2014. Epigenetics of discordant monozygotic twins: implications for disease. *Genome Med,* 6**,** 60.

CHEN, P. Y., FENG, S., JOO, J. W., JACOBSEN, S. E. & PELLEGRINI, M. 2011. A comparative analysis of DNA methylation across human embryonic stem cell lines. *Genome Biol,* 12**,** R62.

CHEN, Y. A., LEMIRE, M., CHOUFANI, S., BUTCHER, D. T., GRAFODATSKAYA, D., ZANKE, B. W., GALLINGER, S., HUDSON, T. J. & WEKSBERG, R. 2013. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics,* 8**,** 203-9.

CITRON, B. A., DENNIS, J. S., ZEITLIN, R. S. & ECHEVERRIA, V. 2008. Transcription factor Sp1 dysregulation in Alzheimer's disease. *J Neurosci Res,* 86**,** 2499-504.

COCHRAN, W. G. & COX, G. M. 1950. *Experimental designs,* New York,, Wiley.

CORDER, E. H., SAUNDERS, A. M., STRITTMATTER, W. J., SCHMECHEL, D. E., GASKELL, P. C., SMALL, G. W., ROSES, A. D., HAINES, J. L. & PERICAK-VANCE, M. A. 1993. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science,* 261**,** 921-3.

COTTON, A. M., PRICE, E. M., JONES, M. J., BALATON, B. P., KOBOR, M. S. & BROWN, C. J. 2015. Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation. *Hum Mol Genet,* 24**,** 1528-39.

CREHAN, H., HOLTON, P., WRAY, S., POCOCK, J., GUERREIRO, R. & HARDY, J. 2012. Complement receptor 1 (CR1) and Alzheimer's disease. *Immunobiology,* 217**,** 244-50.

CREYGHTON, M. P., CHENG, A. W., WELSTEAD, G. G., KOOISTRA, T., CAREY, B. W., STEINE, E. J., HANNA, J., LODATO, M. A., FRAMPTON, G. M., SHARP, P. A., BOYER, L. A., YOUNG, R. A. & JAENISCH, R. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A,* 107**,** 21931-6.

CROISSANT, Y. & MILLO, G. 2008. Panel Data Econometrics in R: The plm Package. *Journal of Statistical Software,* 27.

CRUCHAGA, C., HALLER, G., CHAKRAVERTY, S., MAYO, K., VALLANIA, F. L., MITRA, R. D., FABER, K., WILLIAMSON, J., BIRD, T., DIAZ-ARRASTIA, R., FOROUD, T. M.,

BOEVE, B. F., GRAFF-RADFORD, N. R., ST JEAN, P., LAWSON, M., EHM, M. G., MAYEUX, R., GOATE, A. M. & CONSORTIUM, N.-L. N. F. S. 2012. Rare variants in APP, PSEN1 and PSEN2 increase risk for AD in late-onset Alzheimer's disease families. *PLoS One,* 7**,** e31039.

CUADRADO-TEJEDOR, M., GARCIA-BARROSO, C., SANCHEZ-ARIAS, J. A., RABAL, O., PEREZ-GONZALEZ, M., MEDEROS, S., UGARTE, A., FRANCO, R., SEGURA, V., PEREA, G., OYARZABAL, J. & GARCIA-OSTA, A. 2017. A First-in-Class Small-Molecule that Acts as a Dual Inhibitor of HDAC and PDE5 and that Rescues Hippocampal Synaptic Impairment in Alzheimer's Disease Mice. *Neuropsychopharmacology,* 42**,** 524-539.

DANESE, A., MOFFITT, T. E., ARSENEAULT, L., BLEIBERG, B. A., DINARDO, P. B., GANDELMAN, S. B., HOUTS, R., AMBLER, A., FISHER, H. L., POULTON, R. & CASPI, A. 2016. The Origins of Cognitive Deficits in Victimized Children: Implications for Neuroscientists and Clinicians. *Am J Psychiatry***,** appiajp201616030333.

DANESE, A., MOFFITT, T. E., HARRINGTON, H., MILNE, B. J., POLANCZYK, G., PARIANTE, C. M., POULTON, R. & CASPI, A. 2009. Adverse childhood experiences and adult risk factors for age-related disease: depression, inflammation, and clustering of metabolic risk markers. *Arch Pediatr Adolesc Med,* 163**,** 1135-43.

DANIELS, W. M., FAIRBAIRN, L. R., VAN TILBURG, G., MCEVOY, C. R., ZIGMOND, M. J., RUSSELL, V. A. & STEIN, D. J. 2009. Maternal separation alters nerve growth factor and corticosterone levels but not the DNA methylation status of the exon 1(7) glucocorticoid receptor promoter region. *Metab Brain Dis,* 24**,** 615-27.

DAVEY SMITH, G. & HEMANI, G. 2014. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet,* 23**,** R89-98.

DAVIES, M. N., KRAUSE, L., BELL, J. T., GAO, F., WARD, K. J., WU, H., LU, H., LIU, Y., TSAI, P. C., COLLIER, D. A., MURPHY, T., DEMPSTER, E., MILL, J., CONSORTIUM, U. K. B. E., BATTLE, A., MOSTAFAVI, S., ZHU, X., HENDERS, A., BYRNE, E., WRAY, N. R., MARTIN, N. G., SPECTOR, T. D. & WANG, J. 2014. Hypermethylation in the ZBTB20 gene is associated with major depressive disorder. *Genome Biol,* 15**,** R56.

DAVIES, M. N., VOLTA, M., PIDSLEY, R., LUNNON, K., DIXIT, A., LOVESTONE, S., COARFA, C., HARRIS, R. A., MILOSAVLJEVIC, A., TROAKES, C., AL-SARRAJ, S., DOBSON, R., SCHALKWYK, L. C. & MILL, J. 2012. Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. *Genome Biol,* 13**,** R43.

DAVIES, W., ISLES, A. R. & WILKINSON, L. S. 2005. Imprinted gene expression in the brain. *Neurosci Biobehav Rev,* 29**,** 421-30.

DAVIS, S., BILKE, S., TIM TRICHE, J. & BOOTWALLA, M. 2015. methylumi: Handle Illumina methylation data.

DE JAGER, P. L., SRIVASTAVA, G., LUNNON, K., BURGESS, J., SCHALKWYK, L. C., YU, L., EATON, M. L., KEENAN, B. T., ERNST, J., MCCABE, C., TANG, A., RAJ, T., REPLOGLE, J., BRODEUR, W., GABRIEL, S., CHAI, H. S., YOUNKIN, C., YOUNKIN, S. G., ZOU, F., SZYF, M., EPSTEIN, C. B., SCHNEIDER, J. A., BERNSTEIN, B. E., MEISSNER, A., ERTEKIN-TANER, N., CHIBNIK, L. B., KELLIS, M., MILL, J. & BENNETT, D. A. 2014. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat Neurosci,* 17**,** 1156-63.

DE STROOPER, B., SAFTIG, P., CRAESSAERTS, K., VANDERSTICHELE, H., GUHDE, G., ANNAERT, W., VON FIGURA, K. & VAN LEUVEN, F. 1998. Deficiency of presenilin-1 inhibits the normal cleavage of amyloid precursor protein. *Nature,* 391**,** 387-90.

DEATON, A. M. & BIRD, A. 2011. CpG islands and the regulation of transcription. *Genes Dev,* 25**,** 1010-22.

DEMPSTER, E. L., PIDSLEY, R., SCHALKWYK, L. C., OWENS, S., GEORGIADES, A., KANE, F., KALIDINDI, S., PICCHIONI, M., KRAVARITI, E., TOULOPOULOU, T., MURRAY, R. M. & MILL, J. 2011. Disease-associated epigenetic changes in monozygotic twins discordant for schizophrenia and bipolar disorder. *Hum Mol Genet,* 20**,** 4786-96.

DESARNAUD, F., JAKOVCEVSKI, M., MORELLINI, F. & SCHACHNER, M. 2008. Stress downregulates hippocampal expression of the adhesion molecules NCAM and CHL1 in mice by mechanisms independent of DNA methylation of their promoters. *Cell Adh Migr,* 2**,** 38-44.

DEVLIN, A. M., BRAIN, U., AUSTIN, J. & OBERLANDER, T. F. 2010. Prenatal exposure to maternal depressed mood and the MTHFR C677T variant affect SLC6A4 methylation in infants at birth. *PLoS One,* 5**,** e12201.

DIAS, B. G. & RESSLER, K. J. 2014. Parental olfactory experience influences behavior and neural structure in subsequent generations. *Nat Neurosci,* 17**,** 89-96.

DINGER, M. E., AMARAL, P. P., MERCER, T. R., PANG, K. C., BRUCE, S. J., GARDINER, B. B., ASKARIAN-AMIRI, M. E., RU, K., SOLDA, G., SIMONS, C., SUNKIN, S. M., CROWE, M. L., GRIMMOND, S. M., PERKINS, A. C. & MATTICK, J. S. 2008. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res,* 18**,** 1433-45.

DOHERTY, T. S., FORSTER, A. & ROTH, T. L. 2016. Global and gene-specific DNA methylation alterations in the adolescent amygdala and hippocampus in an animal model of caregiver maltreatment. *Behav Brain Res,* 298**,** 55-61.

DOLINOY, D. C., HUANG, D. & JIRTLE, R. L. 2007. Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. *Proc Natl Acad Sci U S A,* 104**,** 13056-61.

DRINKWATER, R. D., BLAKE, T. J., MORLEY, A. A. & TURNER, D. R. 1989. Human lymphocytes aged in vivo have reduced levels of methylation in transcriptionally active and inactive DNA. *Mutat Res,* 219**,** 29-37.

DUMAN, E. A. & CANLI, T. 2015. Influence of life stress, 5-HTTLPR genotype, and SLC6A4 methylation on gene expression and stress response in healthy Caucasian males. *Biol Mood Anxiety Disord,* 5**,** 2.

ECKERSLEY-MASLIN, M. A., THYBERT, D., BERGMANN, J. H., MARIONI, J. C., FLICEK, P. & SPECTOR, D. L. 2014. Random monoallelic gene expression increases upon embryonic stem cell differentiation. *Dev Cell,* 28**,** 351-65.

ELENKOV, I. J. 2004. Glucocorticoids and the Th1/Th2 balance. *Ann N Y Acad Sci,* 1024**,** 138-46.

ELLIOTT, H. R., TILLIN, T., MCARDLE, W. L., HO, K., DUGGIRALA, A., FRAYLING, T. M., DAVEY SMITH, G., HUGHES, A. D., CHATURVEDI, N. & RELTON, C. L. 2014. Differences in smoking associated DNA methylation patterns in South Asians and Europeans. *Clin Epigenetics,* 6**,** 4.

ENCODE PROJECT CONSORTIUM 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature,* 489**,** 57-74.

ERNST, J. & KELLIS, M. 2010. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol,* 28**,** 817-25.
ERNST, J. & KELLIS, M. 2012. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods,* 9**,** 215-6.
ERNST, J., KHERADPOUR, P., MIKKELSEN, T. S., SHORESH, N., WARD, L. D., EPSTEIN, C. B., ZHANG, X., WANG, L., ISSNER, R., COYNE, M., KU, M., DURHAM, T., KELLIS, M. & BERNSTEIN, B. E. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature,* 473**,** 43-9.
ESPOSITO, E. A., JONES, M. J., DOOM, J. R., MACISAAC, J. L., GUNNAR, M. R. & KOBOR, M. S. 2016. Differential DNA methylation in peripheral blood mononuclear cells in adolescents exposed to significant early but not later childhood adversity. *Dev Psychopathol,* 28**,** 1385-1399.
FEINBERG, A. P. & IRIZARRY, R. A. 2010. Evolution in health and medicine Sackler colloquium: Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc Natl Acad Sci U S A,* 107 Suppl 1**,** 1757-64.
FELSENFELD, G. & GROUDINE, M. 2003. Controlling the double helix. *Nature,* 421**,** 448-53.
FERGUSON-SMITH, A. C. 2011. Genomic imprinting: the emergence of an epigenetic paradigm. *Nat Rev Genet,* 12**,** 565-75.
FIELD, S. F., BERALDI, D., BACHMAN, M., STEWART, S. K., BECK, S. & BALASUBRAMANIAN, S. 2015. Accurate measurement of 5-methylcytosine and 5-hydroxymethylcytosine in human cerebellum DNA by oxidative bisulfite on an array (OxBS-array). *PLoS One,* 10**,** e0118202.
FINKELHOR, D., TURNER, H., HAMBY, S. L. & ORMROD, R. 2011. Polyvictimization: Children's Exposure to Multiple Types of Violence, Crime, and Abuse. *National Survey of Children's Exposure to Violence*.
FISCHER, A. 2014. Targeting histone-modifications in Alzheimer's disease. What is the evidence that this is a promising therapeutic avenue? *Neuropharmacology,* 80**,** 95-102.
FISCHER, A., SANANBENESI, F., WANG, X., DOBBIN, M. & TSAI, L. H. 2007. Recovery of learning and memory is associated with chromatin remodelling. *Nature,* 447**,** 178-82.
FISHER, H. L., CASPI, A., MOFFITT, T. E., WERTZ, J., GRAY, R., NEWBURY, J., AMBLER, A., ZAVOS, H., DANESE, A., MILL, J., ODGERS, C. L., PARIANTE, C., WONG, C. C. & ARSENEAULT, L. 2015. Measuring adolescents' exposure to victimization: The Environmental Risk (E-Risk) Longitudinal Twin Study. *Dev Psychopathol,* 27**,** 1399-416.
FISHER, H. L., MOFFITT, T. E., HOUTS, R. M., BELSKY, D. W., ARSENEAULT, L. & CASPI, A. 2012. Bullying victimisation and risk of self harm in early adolescence: longitudinal cohort study. *BMJ,* 344**,** e2683.
FRAGA, M. F., BALLESTAR, E., PAZ, M. F., ROPERO, S., SETIEN, F., BALLESTAR, M. L., HEINE-SUNER, D., CIGUDOSA, J. C., URIOSTE, M., BENITEZ, J., BOIX-CHORNET, M., SANCHEZ-AGUILERA, A., LING, C., CARLSSON, E., POULSEN, P., VAAG, A., STEPHAN, Z., SPECTOR, T. D., WU, Y. Z., PLASS, C. & ESTELLER, M. 2005. Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci U S A,* 102**,** 10604-9.
FRASER, A., MACDONALD-WALLIS, C., TILLING, K., BOYD, A., GOLDING, J., DAVEY SMITH, G., HENDERSON, J., MACLEOD, J., MOLLOY, L., NESS, A., RING, S.,

NELSON, S. M. & LAWLOR, D. A. 2013. Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol,* 42**,** 97-110.

FUKE, C., SHIMABUKURO, M., PETRONIS, A., SUGIMOTO, J., ODA, T., MIURA, K., MIYAZAKI, T., OGURA, C., OKAZAKI, Y. & JINNO, Y. 2004. Age related changes in 5-methylcytosine content in human peripheral leukocytes and placentas: an HPLC-based study. *Ann Hum Genet,* 68**,** 196-204.

GAL-YAM, E. N., EGGER, G., INIGUEZ, L., HOLSTER, H., EINARSSON, S., ZHANG, X., LIN, J. C., LIANG, G., JONES, P. A. & TANAY, A. 2008. Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. *Proc Natl Acad Sci U S A,* 105**,** 12979-84.

GANDAL, M. J., LEPPA, V., WON, H., PARIKSHAK, N. N. & GESCHWIND, D. H. 2016. The road to precision psychiatry: translating genetics into disease mechanisms. *Nat Neurosci,* 19**,** 1397-1407.

GARDINER-GARDEN, M. & FROMMER, M. 1987. CpG islands in vertebrate genomes. *J Mol Biol,* 196**,** 261-82.

GENDREL, A. V., ATTIA, M., CHEN, C. J., DIABANGOUAYA, P., SERVANT, N., BARILLOT, E. & HEARD, E. 2014. Developmental dynamics and disease potential of random monoallelic gene expression. *Dev Cell,* 28**,** 366-80.

GENTILINI, D., MARI, D., CASTALDI, D., REMONDINI, D., OGLIARI, G., OSTAN, R., BUCCI, L., SIRCHIA, S. M., TABANO, S., CAVAGNINI, F., MONTI, D., FRANCESCHI, C., DI BLASIO, A. M. & VITALE, G. 2013. Role of epigenetics in human aging and longevity: genome-wide DNA methylation profile in centenarians and centenarians' offspring. *Age (Dordr),* 35**,** 1961-73.

GERTZ, J., VARLEY, K. E., REDDY, T. E., BOWLING, K. M., PAULI, F., PARKER, S. L., KUCERA, K. S., WILLARD, H. F. & MYERS, R. M. 2011. Analysis of DNA methylation in a three-generation family reveals widespread genetic influence on epigenetic regulation. *PLoS Genet,* 7**,** e1002228.

GILBERT, R., WIDOM, C. S., BROWNE, K., FERGUSSON, D., WEBB, E. & JANSON, S. 2009. Burden and consequences of child maltreatment in high-income countries. *Lancet,* 373**,** 68-81.

GILFILLAN, G. D., HUGHES, T., SHENG, Y., HJORTHAUG, H. S., STRAUB, T., GERVIN, K., HARRIS, J. R., UNDLIEN, D. E. & LYLE, R. 2012. Limitations and possibilities of low cell number ChIP-seq. *BMC Genomics,* 13**,** 645.

GLOBISCH, D., MUNZEL, M., MULLER, M., MICHALAKIS, S., WAGNER, M., KOCH, S., BRUCKL, T., BIEL, M. & CARELL, T. 2010. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One,* 5**,** e15367.

GLUCKMAN, P. D., HANSON, M. A., COOPER, C. & THORNBURG, K. L. 2008. Effect of in utero and early-life conditions on adult health and disease. *N Engl J Med,* 359**,** 61-73.

GOATE, A., CHARTIER-HARLIN, M. C., MULLAN, M., BROWN, J., CRAWFORD, F., FIDANI, L., GIUFFRA, L., HAYNES, A., IRVING, N., JAMES, L. & ET AL. 1991. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature,* 349**,** 704-6.

GOATE, A. & HARDY, J. 2012. Twenty years of Alzheimer's disease-causing mutations. *J Neurochem,* 120 Suppl 1**,** 3-8.

GOLL, M. G. & BESTOR, T. H. 2005. Eukaryotic cytosine methyltransferases. *Annu Rev Biochem,* 74**,** 481-514.

GRAFF, J. & TSAI, L. H. 2013a. Histone acetylation: molecular mnemonics on the chromatin. *Nat Rev Neurosci,* 14**,** 97-111.
GRAFF, J. & TSAI, L. H. 2013b. The potential of HDAC inhibitors as cognitive enhancers. *Annu Rev Pharmacol Toxicol,* 53**,** 311-30.
GREWAL, S. I. & JIA, S. 2007. Heterochromatin revisited. *Nat Rev Genet,* 8**,** 35-46.
GU, H., SMITH, Z. D., BOCK, C., BOYLE, P., GNIRKE, A. & MEISSNER, A. 2011. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat Protoc,* 6**,** 468-81.
GUAN, J. S., HAGGARTY, S. J., GIACOMETTI, E., DANNENBERG, J. H., JOSEPH, N., GAO, J., NIELAND, T. J., ZHOU, Y., WANG, X., MAZITSCHEK, R., BRADNER, J. E., DEPINHO, R. A., JAENISCH, R. & TSAI, L. H. 2009. HDAC2 negatively regulates memory formation and synaptic plasticity. *Nature,* 459**,** 55-60.
GUILLEMIN, C., PROVENCAL, N., SUDERMAN, M., COTE, S. M., VITARO, F., HALLETT, M., TREMBLAY, R. E. & SZYF, M. 2014. DNA methylation signature of childhood chronic physical aggression in T cells of both men and women. *PLoS One,* 9**,** e86822.
GUINTIVANO, J., ARYEE, M. J. & KAMINSKY, Z. A. 2013. A cell epigenotype specific model for the correction of brain cellular heterogeneity bias and its application to age, brain region and major depression. *Epigenetics,* 8**,** 290-302.
GUNNAR, M. & QUEVEDO, K. 2007. The neurobiology of stress and development. *Annu Rev Psychol,* 58**,** 145-73.
GUTIERREZ-ARCELUS, M., LAPPALAINEN, T., MONTGOMERY, S. B., BUIL, A., ONGEN, H., YUROVSKY, A., BRYOIS, J., GIGER, T., ROMANO, L., PLANCHON, A., FALCONNET, E., BIELSER, D., GAGNEBIN, M., PADIOLEAU, I., BOREL, C., LETOURNEAU, A., MAKRYTHANASIS, P., GUIPPONI, M., GEHRIG, C., ANTONARAKIS, S. E. & DERMITZAKIS, E. T. 2013. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *Elife,* 2**,** e00523.
HAMBY, S. L., FINKELHOR, D., ORMROD, R. K. & TURNER, H. A. 2004. The juvenile victimization questionnaire (JVQ): Administration and scoring manual. *Durham, NH: Crimes Against Children Research Center*.
HANNON, E., DEMPSTER, E., VIANA, J., BURRAGE, J., SMITH, A. R., MACDONALD, R., ST CLAIR, D., MUSTARD, C., BREEN, G., THERMAN, S., KAPRIO, J., TOULOPOULOU, T., HULSHOFF POL, H. E., BOHLKEN, M. M., KAHN, R. S., NENADIC, I., HULTMAN, C. M., MURRAY, R. M., COLLIER, D. A., BASS, N., GURLING, H., MCQUILLIN, A., SCHALKWYK, L. & MILL, J. 2016a. An integrated genetic-epigenetic analysis of schizophrenia: evidence for co-localization of genetic associations and differential DNA methylation. *Genome Biol,* 17**,** 176.
HANNON, E., LUNNON, K., SCHALKWYK, L. & MILL, J. 2015. Interindividual methylomic variation across blood, cortex, and cerebellum: implications for epigenetic studies of neurological and neuropsychiatric phenotypes. *Epigenetics,* 10**,** 1024-32.
HANNON, E., SPIERS, H., VIANA, J., PIDSLEY, R., BURRAGE, J., MURPHY, T. M., TROAKES, C., TURECKI, G., O'DONOVAN, M. C., SCHALKWYK, L. C., BRAY, N. J. & MILL, J. 2016b. Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. *Nat Neurosci,* 19**,** 48-54.
HARDY, J. & SELKOE, D. J. 2002. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science,* 297**,** 353-6.
HARDY, J. A. & HIGGINS, G. A. 1992. Alzheimer's disease: the amyloid cascade hypothesis. *Science,* 256**,** 184-5.

HATA, K., OKANO, M., LEI, H. & LI, E. 2002. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development,* 129**,** 1983-93.

HEARD, E. & MARTIENSSEN, R. A. 2014. Transgenerational epigenetic inheritance: myths and mechanisms. *Cell,* 157**,** 95-109.

HEERBOTH, S., LAPINSKA, K., SNYDER, N., LEARY, M., ROLLINSON, S. & SARKAR, S. 2014. Use of epigenetic drugs in disease: an overview. *Genet Epigenet,* 6**,** 9-19.

HEIJMANS, B. T. & MILL, J. 2012. Commentary: The seven plagues of epigenetic epidemiology. *Int J Epidemiol,* 41**,** 74-8.

HEIJMANS, B. T., TOBI, E. W., STEIN, A. D., PUTTER, H., BLAUW, G. J., SUSSER, E. S., SLAGBOOM, P. E. & LUMEY, L. H. 2008. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A,* 105**,** 17046-9.

HEINZ, S., BENNER, C., SPANN, N., BERTOLINO, E., LIN, Y. C., LASLO, P., CHENG, J. X., MURRE, C., SINGH, H. & GLASS, C. K. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell,* 38**,** 576-89.

HENDRICH, B. & BIRD, A. 1998. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol,* 18**,** 6538-47.

HENNINGSEN, K., DYRVIG, M., BOUZINOVA, E. V., CHRISTIANSEN, S., CHRISTENSEN, T., ANDREASEN, J. T., PALME, R., LICHOTA, J. & WIBORG, O. 2012. Low maternal care exacerbates adult stress susceptibility in the chronic mild stress rat model of depression. *Behav Pharmacol,* 23**,** 735-43.

HEPPNER, F. L., RANSOHOFF, R. M. & BECHER, B. 2015. Immune attack: the role of inflammation in Alzheimer disease. *Nat Rev Neurosci,* 16**,** 358-72.

HEYN, H., LI, N., FERREIRA, H. J., MORAN, S., PISANO, D. G., GOMEZ, A., DIEZ, J., SANCHEZ-MUT, J. V., SETIEN, F., CARMONA, F. J., PUCA, A. A., SAYOLS, S., PUJANA, M. A., SERRA-MUSACH, J., IGLESIAS-PLATAS, I., FORMIGA, F., FERNANDEZ, A. F., FRAGA, M. F., HEATH, S. C., VALENCIA, A., GUT, I. G., WANG, J. & ESTELLER, M. 2012. Distinct DNA methylomes of newborns and centenarians. *Proc Natl Acad Sci U S A,* 109**,** 10522-7.

HORVATH, S. 2013. DNA methylation age of human tissues and cell types. *Genome Biol,* 14**,** R115.

HOUSEMAN, E. A., ACCOMANDO, W. P., KOESTLER, D. C., CHRISTENSEN, B. C., MARSIT, C. J., NELSON, H. H., WIENCKE, J. K. & KELSEY, K. T. 2012. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics,* 13**,** 86.

HOUTEPEN, L. C., VINKERS, C. H., CARRILLO-ROA, T., HIEMSTRA, M., VAN LIER, P. A., MEEUS, W., BRANJE, S., HEIM, C. M., NEMEROFF, C. B., MILL, J., SCHALKWYK, L. C., CREYGHTON, M. P., KAHN, R. S., JOELS, M., BINDER, E. B. & BOKS, M. P. 2016. Genome-wide DNA methylation levels and altered cortisol stress reactivity following childhood trauma in humans. *Nat Commun,* 7**,** 10967.

HUANG, Y., PASTOR, W. A., SHEN, Y., TAHILIANI, M., LIU, D. R. & RAO, A. 2010. The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS One,* 5**,** e8888.

HUEBERT, D. J. & BERNSTEIN, B. E. 2005. Genomic views of chromatin. *Curr Opin Genet Dev,* 15**,** 476-81.

HUISINGA, K. L., BROWER-TOLAND, B. & ELGIN, S. C. 2006. The contradictory definitions of heterochromatin: transcription and silencing. *Chromosoma,* 115**,** 110-22.

HYDE, C. L., NAGLE, M. W., TIAN, C., CHEN, X., PACIGA, S. A., WENDLAND, J. R., TUNG, J. Y., HINDS, D. A., PERLIS, R. H. & WINSLOW, A. R. 2016. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat Genet,* 48**,** 1031-6.

ILLINGWORTH, R. S. & BIRD, A. P. 2009. CpG islands--'a rough guide'. *FEBS Lett,* 583**,** 1713-20.

ILLINGWORTH, R. S., GRUENEWALD-SCHNEIDER, U., WEBB, S., KERR, A. R., JAMES, K. D., TURNER, D. J., SMITH, C., HARRISON, D. J., ANDREWS, R. & BIRD, A. P. 2010. Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet,* 6**,** e1001134.

INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM 2004. Finishing the euchromatic sequence of the human genome. *Nature,* 431**,** 931-45.

IRIZARRY, R. A., LADD-ACOSTA, C., WEN, B., WU, Z., MONTANO, C., ONYANGO, P., CUI, H., GABO, K., RONGIONE, M., WEBSTER, M., JI, H., POTASH, J. B., SABUNCIYAN, S. & FEINBERG, A. P. 2009. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet,* 41**,** 178-86.

ITTNER, L. M. & GOTZ, J. 2011. Amyloid-beta and tau--a toxic pas de deux in Alzheimer's disease. *Nat Rev Neurosci,* 12**,** 65-72.

JAEGER, S. & PIETRZIK, C. U. 2008. Functional role of lipoprotein receptors in Alzheimer's disease. *Curr Alzheimer Res,* 5**,** 15-25.

JAENISCH, R. & BIRD, A. 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet,* 33 Suppl**,** 245-54.

JAFFE, A. E. & IRIZARRY, R. A. 2014. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol,* 15**,** R31.

JAFFEE, S. R., CASPI, A., MOFFITT, T. E. & TAYLOR, A. 2004. Physical maltreatment victim to antisocial child: evidence of an environmentally mediated process. *J Abnorm Psychol,* 113**,** 44-55.

JEFFRIES, A. R. & MILL, J. 2017. Profiling Regulatory Variation in the Brain: Methods for Exploring the Neuronal Epigenome. *Biol Psychiatry,* 81**,** 90-91.

JEFFRIES, A. R., PERFECT, L. W., LEDDEROSE, J., SCHALKWYK, L. C., BRAY, N. J., MILL, J. & PRICE, J. 2012. Stochastic choice of allelic expression in human neural stem cells. *Stem Cells,* 30**,** 1938-47.

JIANG, C. & PUGH, B. F. 2009. Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet,* 10**,** 161-72.

JIN, S. G., KADAM, S. & PFEIFER, G. P. 2010. Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. *Nucleic Acids Res,* 38**,** e125.

JOEHANES, R., JUST, A. C., MARIONI, R. E., PILLING, L. C., REYNOLDS, L. M., MANDAVIYA, P. R., GUAN, W., XU, T., ELKS, C. E., ASLIBEKYAN, S., MORENO-MACIAS, H., SMITH, J. A., BRODY, J. A., DHINGRA, R., YOUSEFI, P., PANKOW, J. S., KUNZE, S., SHAH, S. H., MCRAE, A. F., LOHMAN, K., SHA, J., ABSHER, D. M., FERRUCCI, L., ZHAO, W., DEMERATH, E. W., BRESSLER, J., GROVE, M. L., HUAN, T., LIU, C., MENDELSON, M. M., YAO, C., KIEL, D. P., PETERS, A., WANG-SATTLER, R., VISSCHER, P. M., WRAY, N. R., STARR, J. M., DING, J., RODRIGUEZ,

C. J., WAREHAM, N. J., IRVIN, M. R., ZHI, D., BARRDAHL, M., VINEIS, P., AMBATIPUDI, S., UITTERLINDEN, A. G., HOFMAN, A., SCHWARTZ, J., COLICINO, E., HOU, L., VOKONAS, P. S., HERNANDEZ, D. G., SINGLETON, A. B., BANDINELLI, S., TURNER, S. T., WARE, E. B., SMITH, A. K., KLENGEL, T., BINDER, E. B., PSATY, B. M., TAYLOR, K. D., GHARIB, S. A., SWENSON, B. R., LIANG, L., DEMEO, D. L., O'CONNOR, G. T., HERCEG, Z., RESSLER, K. J., CONNEELY, K. N., SOTOODEHNIA, N., KARDIA, S. L., MELZER, D., BACCARELLI, A. A., VAN MEURS, J. B., ROMIEU, I., ARNETT, D. K., ONG, K. K., LIU, Y., WALDENBERGER, M., DEARY, I. J., FORNAGE, M., LEVY, D. & LONDON, S. J. 2016. Epigenetic Signatures of Cigarette Smoking. *Circ Cardiovasc Genet,* 9**,** 436-447.

JONES, P. A. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet,* 13**,** 484-92.

JONSSON, T., ATWAL, J. K., STEINBERG, S., SNAEDAL, J., JONSSON, P. V., BJORNSSON, S., STEFANSSON, H., SULEM, P., GUDBJARTSSON, D., MALONEY, J., HOYTE, K., GUSTAFSON, A., LIU, Y., LU, Y., BHANGALE, T., GRAHAM, R. R., HUTTENLOCHER, J., BJORNSDOTTIR, G., ANDREASSEN, O. A., JONSSON, E. G., PALOTIE, A., BEHRENS, T. W., MAGNUSSON, O. T., KONG, A., THORSTEINSDOTTIR, U., WATTS, R. J. & STEFANSSON, K. 2012. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature,* 488**,** 96-9.

KAAS, G. A., ZHONG, C., EASON, D. E., ROSS, D. L., VACHHANI, R. V., MING, G. L., KING, J. R., SONG, H. & SWEATT, J. D. 2013. TET1 controls CNS 5-methylcytosine hydroxylation, active DNA demethylation, gene transcription, and memory formation. *Neuron,* 79**,** 1086-93.

KAIKKONEN, M. U., LAM, M. T. & GLASS, C. K. 2011. Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovasc Res,* 90**,** 430-40.

KAMINSKY, Z., WANG, S. C. & PETRONIS, A. 2006. Complex disease, gender and epigenetics. *Ann Med,* 38**,** 530-44.

KAMINSKY, Z. A., TANG, T., WANG, S. C., PTAK, C., OH, G. H., WONG, A. H., FELDCAMP, L. A., VIRTANEN, C., HALFVARSON, J., TYSK, C., MCRAE, A. F., VISSCHER, P. M., MONTGOMERY, G. W., GOTTESMAN, II, MARTIN, N. G. & PETRONIS, A. 2009. DNA methylation profiles in monozygotic and dizygotic twins. *Nat Genet,* 41**,** 240-5.

KAPRANOV, P., CHENG, J., DIKE, S., NIX, D. A., DUTTAGUPTA, R., WILLINGHAM, A. T., STADLER, P. F., HERTEL, J., HACKERMULLER, J., HOFACKER, I. L., BELL, I., CHEUNG, E., DRENKOW, J., DUMAIS, E., PATEL, S., HELT, G., GANESH, M., GHOSH, S., PICCOLBONI, A., SEMENTCHENKO, V., TAMMANA, H. & GINGERAS, T. R. 2007. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science,* 316**,** 1484-8.

KARCH, C. M., CRUCHAGA, C. & GOATE, A. M. 2014. Alzheimer's disease genetics: from the bench to the clinic. *Neuron,* 83**,** 11-26.

KASOWSKI, M., KYRIAZOPOULOU-PANAGIOTOPOULOU, S., GRUBERT, F., ZAUGG, J. B., KUNDAJE, A., LIU, Y., BOYLE, A. P., ZHANG, Q. C., ZAKHARIA, F., SPACEK, D. V., LI, J., XIE, D., OLARERIN-GEORGE, A., STEINMETZ, L. M., HOGENESCH, J. B., KELLIS, M., BATZOGLOU, S. & SNYDER, M. 2013. Extensive variation in chromatin states across humans. *Science,* 342**,** 750-2.

KEMBER, R. L., DEMPSTER, E. L., LEE, T. H., SCHALKWYK, L. C., MILL, J. & FERNANDES, C. 2012. Maternal separation is associated with strain-specific responses to stress and epigenetic alterations to Nr3c1, Avp, and Nr4a1 in mouse. *Brain Behav,* 2**,** 455-67.

KERKEL, K., SPADOLA, A., YUAN, E., KOSEK, J., JIANG, L., HOD, E., LI, K., MURTY, V. V., SCHUPF, N., VILAIN, E., MORRIS, M., HAGHIGHI, F. & TYCKO, B. 2008. Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat Genet,* 40**,** 904-8.

KILGORE, M., MILLER, C. A., FASS, D. M., HENNIG, K. M., HAGGARTY, S. J., SWEATT, J. D. & RUMBAUGH, G. 2010. Inhibitors of class 1 histone deacetylases reverse contextual memory deficits in a mouse model of Alzheimer's disease. *Neuropsychopharmacology,* 35**,** 870-80.

KILPINEN, H., WASZAK, S. M., GSCHWIND, A. R., RAGHAV, S. K., WITWICKI, R. M., ORIOLI, A., MIGLIAVACCA, E., WIEDERKEHR, M., GUTIERREZ-ARCELUS, M., PANOUSIS, N. I., YUROVSKY, A., LAPPALAINEN, T., ROMANO-PALUMBO, L., PLANCHON, A., BIELSER, D., BRYOIS, J., PADIOLEAU, I., UDIN, G., THURNHEER, S., HACKER, D., CORE, L. J., LIS, J. T., HERNANDEZ, N., REYMOND, A., DEPLANCKE, B. & DERMITZAKIS, E. T. 2013. Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. *Science,* 342**,** 744-7.

KINDE, B., GABEL, H. W., GILBERT, C. S., GRIFFITH, E. C. & GREENBERG, M. E. 2015. Reading the unique DNA methylation landscape of the brain: Non-CpG methylation, hydroxymethylation, and MeCP2. *Proc Natl Acad Sci U S A,* 112**,** 6800-6.

KINNALLY, E. L., CAPITANIO, J. P., LEIBEL, R., DENG, L., LEDUC, C., HAGHIGHI, F. & MANN, J. J. 2010. Epigenetic regulation of serotonin transporter expression and behavior in infant rhesus macaques. *Genes Brain Behav,* 9**,** 575-82.

KLENGEL, T., MEHTA, D., ANACKER, C., REX-HAFFNER, M., PRUESSNER, J. C., PARIANTE, C. M., PACE, T. W., MERCER, K. B., MAYBERG, H. S., BRADLEY, B., NEMEROFF, C. B., HOLSBOER, F., HEIM, C. M., RESSLER, K. J., REIN, T. & BINDER, E. B. 2013. Allele-specific FKBP5 DNA demethylation mediates gene-childhood trauma interactions. *Nat Neurosci,* 16**,** 33-41.

KLOSE, R. J. & BIRD, A. P. 2006. Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci,* 31**,** 89-97.

KNAPP, M., PRINCE, M., ALBANESE, E., BANERJEE, S., DHANASIRI, S. & FERNANDEZ, J. 2007. Dementia UK: The full report. London: Alzheimer's Society.

KOHLI, R. M. & ZHANG, Y. 2013. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature,* 502**,** 472-9.

KONG, A., STEINTHORSDOTTIR, V., MASSON, G., THORLEIFSSON, G., SULEM, P., BESENBACHER, S., JONASDOTTIR, A., SIGURDSSON, A., KRISTINSSON, K. T., JONASDOTTIR, A., FRIGGE, M. L., GYLFASON, A., OLASON, P. I., GUDJONSSON, S. A., SVERRISSON, S., STACEY, S. N., SIGURGEIRSSON, B., BENEDIKTSDOTTIR, K. R., SIGURDSSON, H., JONSSON, T., BENEDIKTSSON, R., OLAFSSON, J. H., JOHANNSSON, O. T., HREIDARSSON, A. B., SIGURDSSON, G., CONSORTIUM, D., FERGUSON-SMITH, A. C., GUDBJARTSSON, D. F., THORSTEINSDOTTIR, U. & STEFANSSON, K. 2009. Parental origin of sequence variants associated with complex diseases. *Nature,* 462**,** 868-74.

KOSTEN, T. A., HUANG, W. & NIELSEN, D. A. 2014. Sex and litter effects on anxiety and DNA methylation levels of stress and neurotrophin genes in adolescent rats. *Dev Psychobiol,* 56**,** 392-406.

KOSTEN, T. A. & NIELSEN, D. A. 2014. Litter and sex effects on maternal behavior and DNA methylation of the Nr3c1 exon 17 promoter gene in hippocampus and cerebellum. *Int J Dev Neurosci,* 36**,** 5-12.

KOZLENKOV, A., ROUSSOS, P., TIMASHPOLSKY, A., BARBU, M., RUDCHENKO, S., BIBIKOVA, M., KLOTZLE, B., BYNE, W., LYDDON, R., DI NARZO, A. F., HURD, Y. L., KOONIN, E. V. & DRACHEVA, S. 2014. Differences in DNA methylation between human neuronal and glial cells are concentrated in enhancers and non-CpG sites. *Nucleic Acids Res,* 42**,** 109-27.

KRETZSCHMAR, H. 2009. Brain banking: opportunities, challenges and meaning for the future. *Nat Rev Neurosci,* 10**,** 70-8.

KRIAUCIONIS, S. & HEINTZ, N. 2009. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science,* 324**,** 929-30.

KROL, J., LOEDIGE, I. & FILIPOWICZ, W. 2010. The widespread regulation of microRNA biogenesis, function and decay. *Nat Rev Genet,* 11**,** 597-610.

KUMSTA, R., KREPPNER, J., KENNEDY, M., KNIGHTS, N., RUTTER, M. & SONUGA-BARKE, E. 2015. Psychological Consequences of Early Global Deprivation An Overview of Findings From the English & Romanian Adoptees Study. *European Psychologist,* 20**,** 138-151.

KUMSTA, R., MARZI, S. J., VIANA, J., DEMPSTER, E. L., CRAWFORD, B., RUTTER, M., MILL, J. & SONUGA-BARKE, E. J. 2016. Severe psychosocial deprivation in early childhood is associated with increased DNA methylation across a region spanning the transcription start site of CYP2E1. *Transl Psychiatry,* 6**,** e830.

KUNDAJE, A., MEULEMAN, W., ERNST, J., BILENKY, M., YEN, A., HERAVI-MOUSSAVI, A., KHERADPOUR, P., ZHANG, Z., WANG, J., ZILLER, M. J., AMIN, V., WHITAKER, J. W., SCHULTZ, M. D., WARD, L. D., SARKAR, A., QUON, G., SANDSTROM, R. S., EATON, M. L., WU, Y. C., PFENNING, A. R., WANG, X., CLAUSSNITZER, M., LIU, Y., COARFA, C., HARRIS, R. A., SHORESH, N., EPSTEIN, C. B., GJONESKA, E., LEUNG, D., XIE, W., HAWKINS, R. D., LISTER, R., HONG, C., GASCARD, P., MUNGALL, A. J., MOORE, R., CHUAH, E., TAM, A., CANFIELD, T. K., HANSEN, R. S., KAUL, R., SABO, P. J., BANSAL, M. S., CARLES, A., DIXON, J. R., FARH, K. H., FEIZI, S., KARLIC, R., KIM, A. R., KULKARNI, A., LI, D., LOWDON, R., ELLIOTT, G., MERCER, T. R., NEPH, S. J., ONUCHIC, V., POLAK, P., RAJAGOPAL, N., RAY, P., SALLARI, R. C., SIEBENTHALL, K. T., SINNOTT-ARMSTRONG, N. A., STEVENS, M., THURMAN, R. E., WU, J., ZHANG, B., ZHOU, X., BEAUDET, A. E., BOYER, L. A., DE JAGER, P. L., FARNHAM, P. J., FISHER, S. J., HAUSSLER, D., JONES, S. J., LI, W., MARRA, M. A., MCMANUS, M. T., SUNYAEV, S., THOMSON, J. A., TLSTY, T. D., TSAI, L. H., WANG, W., WATERLAND, R. A., ZHANG, M. Q., CHADWICK, L. H., BERNSTEIN, B. E., COSTELLO, J. F., ECKER, J. R., HIRST, M., MEISSNER, A., MILOSAVLJEVIC, A., REN, B., STAMATOYANNOPOULOS, J. A., WANG, T. & KELLIS, M. 2015. Integrative analysis of 111 reference human epigenomes. *Nature,* 518**,** 317-30.

KUNDAKOVIC, M., LIM, S., GUDSNUK, K. & CHAMPAGNE, F. A. 2013. Sex-specific and strain-dependent effects of early life adversity on behavioral and epigenetic outcomes. *Front Psychiatry,* 4**,** 78.

LABONTE, B., YERKO, V., GROSS, J., MECHAWAR, N., MEANEY, M. J., SZYF, M. & TURECKI, G. 2012. Differential glucocorticoid receptor exon 1(B), 1(C), and 1(H) expression and methylation in suicide completers with a history of childhood abuse. *Biol Psychiatry,* 72**,** 41-8.

LADD-ACOSTA, C., PEVSNER, J., SABUNCIYAN, S., YOLKEN, R. H., WEBSTER, M. J., DINKINS, T., CALLINAN, P. A., FAN, J. B., POTASH, J. B. & FEINBERG, A. P. 2007. DNA methylation signatures within the human brain. *Am J Hum Genet,* 81**,** 1304-15.

LAIRD, P. W. 2010. Principles and challenges of genomewide DNA methylation analysis. *Nat Rev Genet,* 11**,** 191-203.

LAMBERT, J. C., IBRAHIM-VERBAAS, C. A., HAROLD, D., NAJ, A. C., SIMS, R., BELLENGUEZ, C., DESTAFANO, A. L., BIS, J. C., BEECHAM, G. W., GRENIER-BOLEY, B., RUSSO, G., THORTON-WELLS, T. A., JONES, N., SMITH, A. V., CHOURAKI, V., THOMAS, C., IKRAM, M. A., ZELENIKA, D., VARDARAJAN, B. N., KAMATANI, Y., LIN, C. F., GERRISH, A., SCHMIDT, H., KUNKLE, B., DUNSTAN, M. L., RUIZ, A., BIHOREAU, M. T., CHOI, S. H., REITZ, C., PASQUIER, F., CRUCHAGA, C., CRAIG, D., AMIN, N., BERR, C., LOPEZ, O. L., DE JAGER, P. L., DERAMECOURT, V., JOHNSTON, J. A., EVANS, D., LOVESTONE, S., LETENNEUR, L., MORON, F. J., RUBINSZTEIN, D. C., EIRIKSDOTTIR, G., SLEEGERS, K., GOATE, A. M., FIEVET, N., HUENTELMAN, M. W., GILL, M., BROWN, K., KAMBOH, M. I., KELLER, L., BARBERGER-GATEAU, P., MCGUINESS, B., LARSON, E. B., GREEN, R., MYERS, A. J., DUFOUIL, C., TODD, S., WALLON, D., LOVE, S., ROGAEVA, E., GALLACHER, J., ST GEORGE-HYSLOP, P., CLARIMON, J., LLEO, A., BAYER, A., TSUANG, D. W., YU, L., TSOLAKI, M., BOSSU, P., SPALLETTA, G., PROITSI, P., COLLINGE, J., SORBI, S., SANCHEZ-GARCIA, F., FOX, N. C., HARDY, J., DENIZ NARANJO, M. C., BOSCO, P., CLARKE, R., BRAYNE, C., GALIMBERTI, D., MANCUSO, M., MATTHEWS, F., EUROPEAN ALZHEIMER'S DISEASE, I., GENETIC, ENVIRONMENTAL RISK IN ALZHEIMER'S, D., ALZHEIMER'S DISEASE GENETIC, C., COHORTS FOR, H., AGING RESEARCH IN GENOMIC, E., MOEBUS, S., MECOCCI, P., DEL ZOMPO, M., MAIER, W., HAMPEL, H., PILOTTO, A., BULLIDO, M., PANZA, F., CAFFARRA, P., et al. 2013. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet,* 45**,** 1452-8.

LANGMEAD, B., TRAPNELL, C., POP, M. & SALZBERG, S. L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol,* 10**,** R25.

LAW, J. A. & JACOBSEN, S. E. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet,* 11**,** 204-20.

LAWRENCE, M., HUBER, W., PAGES, H., ABOYOUN, P., CARLSON, M., GENTLEMAN, R., MORGAN, M. T. & CAREY, V. J. 2013. Software for computing and annotating genomic ranges. *PLoS Comput Biol,* 9**,** e1003118.

LEE, J. T. & BARTOLOMEI, M. S. 2013. X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell,* 152**,** 1308-23.

LEEK, J. T., SCHARPF, R. B., BRAVO, H. C., SIMCHA, D., LANGMEAD, B., JOHNSON, W. E., GEMAN, D., BAGGERLY, K. & IRIZARRY, R. A. 2010. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet,* 11**,** 733-9.

LI, E., BESTOR, T. H. & JAENISCH, R. 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell,* 69**,** 915-26.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA PROCESSING, S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.

LIBERMAN, S. A., MASHOODH, R., THOMPSON, R. C., DOLINOY, D. C. & CHAMPAGNE, F. A. 2012. Concordance in hippocampal and fecal Nr3c1 methylation is moderated by maternal behavior in the mouse. *Ecol Evol,* 2**,** 3123-31.

LIMON, A., REYES-RUIZ, J. M. & MILEDI, R. 2012. Loss of functional GABA(A) receptors in the Alzheimer diseased brain. *Proc Natl Acad Sci U S A,* 109**,** 10071-6.

LISTER, R., MUKAMEL, E. A., NERY, J. R., URICH, M., PUDDIFOOT, C. A., JOHNSON, N. D., LUCERO, J., HUANG, Y., DWORK, A. J., SCHULTZ, M. D., YU, M., TONTI-FILIPPINI, J., HEYN, H., HU, S., WU, J. C., RAO, A., ESTELLER, M., HE, C., HAGHIGHI, F. G., SEJNOWSKI, T. J., BEHRENS, M. M. & ECKER, J. R. 2013. Global epigenomic reconfiguration during mammalian brain development. *Science,* 341**,** 1237905.

LISTER, R., PELIZZOLA, M., DOWEN, R. H., HAWKINS, R. D., HON, G., TONTI-FILIPPINI, J., NERY, J. R., LEE, L., YE, Z., NGO, Q. M., EDSALL, L., ANTOSIEWICZ-BOURGET, J., STEWART, R., RUOTTI, V., MILLAR, A. H., THOMSON, J. A., REN, B. & ECKER, J. R. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature,* 462**,** 315-22.

LIU, C. C., KANEKIYO, T., XU, H. & BU, G. 2013. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nat Rev Neurol,* 9**,** 106-18.

LIU, D., DIORIO, J., TANNENBAUM, B., CALDJI, C., FRANCIS, D., FREEDMAN, A., SHARMA, S., PEARSON, D., PLOTSKY, P. M. & MEANEY, M. J. 1997. Maternal care, hippocampal glucocorticoid receptors, and hypothalamic-pituitary-adrenal responses to stress. *Science,* 277**,** 1659-62.

LOVE, M. I., HUBER, W. & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol,* 15**,** 550.

LOWE, R., GEMMA, C., BEYAN, H., HAWA, M. I., BAZEOS, A., LESLIE, R. D., MONTPETIT, A., RAKYAN, V. K. & RAMAGOPALAN, S. V. 2013. Buccals are likely to be a more informative surrogate tissue than blood for epigenome-wide association studies. *Epigenetics,* 8**,** 445-54.

LU, T., ARON, L., ZULLO, J., PAN, Y., KIM, H., CHEN, Y., YANG, T. H., KIM, H. M., DRAKE, D., LIU, X. S., BENNETT, D. A., COLAIACOVO, M. P. & YANKNER, B. A. 2014. REST and stress resistance in ageing and Alzheimer's disease. *Nature,* 507**,** 448-54.

LU, X., WANG, L., YU, C., YU, D. & YU, G. 2015. Histone Acetylation Modifiers in the Pathogenesis of Alzheimer's Disease. *Front Cell Neurosci,* 9**,** 226.

LUGER, K., MADER, A. W., RICHMOND, R. K., SARGENT, D. F. & RICHMOND, T. J. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature,* 389**,** 251-60.

LUMEY, L. H., STEIN, A. D., KAHN, H. S., VAN DER PAL-DE BRUIN, K. M., BLAUW, G. J., ZYBERT, P. A. & SUSSER, E. S. 2007. Cohort profile: the Dutch Hunger Winter families study. *Int J Epidemiol,* 36**,** 1196-204.

LUN, A. T., CHEN, Y. & SMYTH, G. K. 2016. It's DE-licious: A Recipe for Differential Expression Analyses of RNA-seq Experiments Using Quasi-Likelihood Methods in edgeR. *Methods Mol Biol,* 1418**,** 391-416.

LUNNON, K., HANNON, E., SMITH, R. G., DEMPSTER, E., WONG, C., BURRAGE, J., TROAKES, C., AL-SARRAJ, S., KEPA, A., SCHALKWYK, L. & MILL, J. 2016. Variation in 5-hydroxymethylcytosine across human cortex and cerebellum. *Genome Biol,* 17**,** 27.

LUNNON, K. & MILL, J. 2013. Epigenetic studies in Alzheimer's disease: current findings, caveats, and considerations for future studies. *Am J Med Genet B Neuropsychiatr Genet,* 162B**,** 789-99.

LUNNON, K., SMITH, R., HANNON, E., DE JAGER, P. L., SRIVASTAVA, G., VOLTA, M., TROAKES, C., AL-SARRAJ, S., BURRAGE, J., MACDONALD, R., CONDLIFFE, D., HARRIES, L. W., KATSEL, P., HAROUTUNIAN, V., KAMINSKY, Z., JOACHIM, C., POWELL, J., LOVESTONE, S., BENNETT, D. A., SCHALKWYK, L. C. & MILL, J. 2014. Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nat Neurosci,* 17**,** 1164-70.

LUPIEN, S. J., MCEWEN, B. S., GUNNAR, M. R. & HEIM, C. 2009. Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nat Rev Neurosci,* 10**,** 434-45.

LUTZ, P. E. & TURECKI, G. 2014. DNA methylation and childhood maltreatment: from animal models to human studies. *Neuroscience,* 264**,** 142-56.

MARIONI, R. E., SHAH, S., MCRAE, A. F., CHEN, B. H., COLICINO, E., HARRIS, S. E., GIBSON, J., HENDERS, A. K., REDMOND, P., COX, S. R., PATTIE, A., CORLEY, J., MURPHY, L., MARTIN, N. G., MONTGOMERY, G. W., FEINBERG, A. P., FALLIN, M. D., MULTHAUP, M. L., JAFFE, A. E., JOEHANES, R., SCHWARTZ, J., JUST, A. C., LUNETTA, K. L., MURABITO, J. M., STARR, J. M., HORVATH, S., BACCARELLI, A. A., LEVY, D., VISSCHER, P. M., WRAY, N. R. & DEARY, I. J. 2015. DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol,* 16**,** 25.

MARQUAND, A. F., REZEK, I., BUITELAAR, J. & BECKMANN, C. F. 2016. Understanding Heterogeneity in Clinical Cohorts Using Normative Models: Beyond Case-Control Studies. *Biol Psychiatry,* 80**,** 552-61.

MARTIN-BLANCO, A., FERRER, M., SOLER, J., SALAZAR, J., VEGA, D., ANDION, O., SANCHEZ-MORA, C., ARRANZ, M. J., RIBASES, M., FELIU-SOLER, A., PEREZ, V. & PASCUAL, J. C. 2014. Association between methylation of the glucocorticoid receptor gene, childhood maltreatment, and clinical severity in borderline personality disorder. *J Psychiatr Res,* 57**,** 34-40.

MARZI, S. J., MEABURN, E. L., DEMPSTER, E. L., LUNNON, K., PAYA-CANO, J. L., SMITH, R. G., VOLTA, M., TROAKES, C., SCHALKWYK, L. C. & MILL, J. 2016. Tissue-specific patterns of allelically-skewed DNA methylation. *Epigenetics,* 11**,** 24-35.

MATHELIER, A., FORNES, O., ARENILLAS, D. J., CHEN, C. Y., DENAY, G., LEE, J., SHI, W., SHYR, C., TAN, G., WORSLEY-HUNT, R., ZHANG, A. W., PARCY, F., LENHARD, B., SANDELIN, A. & WASSERMAN, W. W. 2016. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res,* 44**,** D110-5.

MAUNAKEA, A. K., NAGARAJAN, R. P., BILENKY, M., BALLINGER, T. J., D'SOUZA, C., FOUSE, S. D., JOHNSON, B. E., HONG, C., NIELSEN, C., ZHAO, Y., TURECKI, G., DELANEY, A., VARHOL, R., THIESSEN, N., SHCHORS, K., HEINE, V. M., ROWITCH, D. H., XING, X., FIORE, C., SCHILLEBEECKX, M., JONES, S. J., HAUSSLER, D., MARRA, M. A., HIRST, M., WANG, T. & COSTELLO, J. F. 2010. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature,* 466**,** 253-7.

MAURANO, M. T., HUMBERT, R., RYNES, E., THURMAN, R. E., HAUGEN, E., WANG, H., REYNOLDS, A. P., SANDSTROM, R., QU, H., BRODY, J., SHAFER, A., NERI, F., LEE, K., KUTYAVIN, T., STEHLING-SUN, S., JOHNSON, A. K., CANFIELD, T. K., GISTE, E., DIEGEL, M., BATES, D., HANSEN, R. S., NEPH, S., SABO, P. J., HEIMFELD, S., RAUBITSCHEK, A., ZIEGLER, S., COTSAPAS, C., SOTOODEHNIA, N., GLASS, I., SUNYAEV, S. R., KAUL, R. & STAMATOYANNOPOULOS, J. A. 2012. Systematic localization of common disease-associated variation in regulatory DNA. *Science,* 337**,** 1190-5.

MCEWEN, B. S. 2012. Brain on stress: how the social environment gets under the skin. *Proc Natl Acad Sci U S A,* 109 Suppl 2**,** 17180-5.

MCGOWAN, P. O., SASAKI, A., D'ALESSIO, A. C., DYMOV, S., LABONTE, B., SZYF, M., TURECKI, G. & MEANEY, M. J. 2009. Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nat Neurosci,* 12**,** 342-8.

MCLEAN, C. Y., BRISTOR, D., HILLER, M., CLARKE, S. L., SCHAAR, B. T., LOWE, C. B., WENGER, A. M. & BEJERANO, G. 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol,* 28**,** 495-501.

MCVICKER, G., VAN DE GEIJN, B., DEGNER, J. F., CAIN, C. E., BANOVICH, N. E., RAJ, A., LEWELLEN, N., MYRTHIL, M., GILAD, Y. & PRITCHARD, J. K. 2013. Identification of genetic variants that affect histone modifications in human cells. *Science,* 342**,** 747-9.

MEABURN, E. L., SCHALKWYK, L. C. & MILL, J. 2010. Allele-specific methylation in the human genome: implications for genetic studies of complex disease. *Epigenetics,* 5**,** 578-82.

MEHTA, D., KLENGEL, T., CONNEELY, K. N., SMITH, A. K., ALTMANN, A., PACE, T. W., REX-HAFFNER, M., LOESCHNER, A., GONIK, M., MERCER, K. B., BRADLEY, B., MULLER-MYHSOK, B., RESSLER, K. J. & BINDER, E. B. 2013. Childhood maltreatment is associated with distinct genomic and epigenetic profiles in posttraumatic stress disorder. *Proc Natl Acad Sci U S A,* 110**,** 8302-7.

MEISSNER, A., GNIRKE, A., BELL, G. W., RAMSAHOYE, B., LANDER, E. S. & JAENISCH, R. 2005. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res,* 33**,** 5868-77.

MELAS, P. A., WEI, Y., WONG, C. C., SJOHOLM, L. K., ABERG, E., MILL, J., SCHALLING, M., FORSELL, Y. & LAVEBRATT, C. 2013. Genetic and epigenetic associations of MAOA and NR3C1 with depression and childhood adversities. *Int J Neuropsychopharmacol,* 16**,** 1513-28.

MENKE, A. & BINDER, E. B. 2014. Epigenetic alterations in depression and antidepressant treatment. *Dialogues Clin Neurosci,* 16**,** 395-404.

MERCER, T. R., DINGER, M. E. & MATTICK, J. S. 2009. Long non-coding RNAs: insights into functions. *Nat Rev Genet,* 10**,** 155-9.

MEYER-LINDENBERG, A., DOMES, G., KIRSCH, P. & HEINRICHS, M. 2011. Oxytocin and vasopressin in the human brain: social neuropeptides for translational medicine. *Nat Rev Neurosci,* 12**,** 524-38.

MILEVA, G., BAKER, S. L., KONKLE, A. T. & BIELAJEW, C. 2014. Bisphenol-A: epigenetic reprogramming and effects on reproduction and behavior. *Int J Environ Res Public Health,* 11**,** 7537-61.

MILL, J. & HEIJMANS, B. T. 2013. From promises to practical strategies in epigenetic epidemiology. *Nat Rev Genet,* 14**,** 585-94.

MO, A., MUKAMEL, E. A., DAVIS, F. P., LUO, C., HENRY, G. L., PICARD, S., URICH, M. A., NERY, J. R., SEJNOWSKI, T. J., LISTER, R., EDDY, S. R., ECKER, J. R. & NATHANS, J. 2015. Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. *Neuron,* 86**,** 1369-84.

MOFFITT, T. E. & E-RISK STUDY TEAM 2002. Teen-aged mothers in contemporary Britain. *J Child Psychol Psychiatry,* 43**,** 727-42.

MURGATROYD, C., PATCHEV, A. V., WU, Y., MICALE, V., BOCKMUHL, Y., FISCHER, D., HOLSBOER, F., WOTJAK, C. T., ALMEIDA, O. F. & SPENGLER, D. 2009. Dynamic DNA methylation programs persistent adverse effects of early-life stress. *Nat Neurosci,* 12**,** 1559-66.

MURPHY, T. M., CRAWFORD, B., DEMPSTER, E. L., HANNON, E., BURRAGE, J., TURECKI, G., KAMINSKY, Z. & MILL, J. 2017. Methylomic profiling of cortex samples from completed suicide cases implicates a role for PSORS1C3 in major depression and suicide. *Transl Psychiatry,* 7**,** e989.

NARAYAN, P. J., LILL, C., FAULL, R., CURTIS, M. A. & DRAGUNOW, M. 2015. Increased acetyl and total histone levels in post-mortem Alzheimer's disease brain. *Neurobiol Dis,* 74**,** 281-94.

NAUMOVA, O. Y., LEE, M., KOPOSOV, R., SZYF, M., DOZIER, M. & GRIGORENKO, E. L. 2012. Differential patterns of whole-genome DNA methylation in institutionalized children and children raised by their biological parents. *Dev Psychopathol,* 24**,** 143-55.

NICHOLS, H. B. & HARLOW, B. L. 2004. Childhood abuse and risk of smoking onset. *J Epidemiol Community Health,* 58**,** 402-6.

NICOLAE, D. L., GAMAZON, E., ZHANG, W., DUAN, S., DOLAN, M. E. & COX, N. J. 2010. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet,* 6**,** e1000888.

NON, A. L., HOLLISTER, B. M., HUMPHREYS, K. L., CHILDEBAYEVA, A., ESTEVES, K., ZEANAH, C. H., FOX, N. A., NELSON, C. A. & DRURY, S. S. 2016. DNA methylation at stress-related genes is associated with exposure to early life institutionalization. *Am J Phys Anthropol,* 161**,** 84-93.

NORMAN, R. E., BYAMBAA, M., DE, R., BUTCHART, A., SCOTT, J. & VOS, T. 2012. The long-term health consequences of child physical abuse, emotional abuse, and neglect: a systematic review and meta-analysis. *PLoS Med,* 9**,** e1001349.

O'NEILL, L. P. & TURNER, B. M. 2003. Immunoprecipitation of native chromatin: NChIP. *Methods,* 31**,** 76-82.

ODGERS, C. L., CASPI, A., BATES, C. J., SAMPSON, R. J. & MOFFITT, T. E. 2012. Systematic social observation of children's neighborhoods using Google Street View: a reliable and cost-effective method. *J Child Psychol Psychiatry,* 53**,** 1009-17.

OKANO, M., BELL, D. W., HABER, D. A. & LI, E. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell,* 99**,** 247-57.

OOI, S. K., QIU, C., BERNSTEIN, E., LI, K., JIA, D., YANG, Z., ERDJUMENT-BROMAGE, H., TEMPST, P., LIN, S. P., ALLIS, C. D., CHENG, X. & BESTOR, T. H. 2007. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature,* 448**,** 714-7.

OSBORNE, J. D., FLATOW, J., HOLKO, M., LIN, S. M., KIBBE, W. A., ZHU, L. J., DANILA, M. I., FENG, G. & CHISHOLM, R. L. 2009. Annotating the human genome with Disease Ontology. *BMC Genomics,* 10 Suppl 1**,** S6.

OUELLET-MORIN, I., WONG, C. C., DANESE, A., PARIANTE, C. M., PAPADOPOULOS, A. S., MILL, J. & ARSENEAULT, L. 2013. Increased serotonin transporter gene (SERT) DNA methylation is associated with bullying victimization and blunted cortisol response to stress in childhood: a longitudinal study of discordant monozygotic twins. *Psychol Med,* 43**,** 1813-23.

PALACIOS, R. & SUGAWARA, I. 1982. Hydrocortisone abrogates proliferation of T cells in autologous mixed lymphocyte reaction by rendering the interleukin-2 Producer T cells unresponsive to interleukin-1 and unable to synthesize the T-cell growth factor. *Scand J Immunol,* 15**,** 25-31.

PALMA-GUDIEL, H., CORDOVA-PALOMERA, A., LEZA, J. C. & FANANAS, L. 2015. Glucocorticoid receptor gene (NR3C1) methylation processes as mediators of early adversity in stress-related disorders causality: A critical review. *Neurosci Biobehav Rev,* 55**,** 520-35.

PESKIN, M. F., TORTOLERO, S. R. & MARKHAM, C. M. 2006. Bullying and victimization among black and Hispanic adolescents. *Adolescence,* 41**,** 467-84.
PETRONIS, A. 2010. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature,* 465**,** 721-7.
PIDSLEY, R., CC, Y. W., VOLTA, M., LUNNON, K., MILL, J. & SCHALKWYK, L. C. 2013. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics,* 14**,** 293.
PIDSLEY, R., VIANA, J., HANNON, E., SPIERS, H., TROAKES, C., AL-SARAJ, S., MECHAWAR, N., TURECKI, G., SCHALKWYK, L. C., BRAY, N. J. & MILL, J. 2014. Methylomic profiling of human brain tissue supports a neurodevelopmental origin for schizophrenia. *Genome Biol,* 15**,** 483.
POLANCZYK, G., CASPI, A., WILLIAMS, B., PRICE, T. S., DANESE, A., SUGDEN, K., UHER, R., POULTON, R. & MOFFITT, T. E. 2009. Protective effect of CRHR1 gene variants on the development of adult depression following childhood maltreatment: replication and extension. *Arch Gen Psychiatry,* 66**,** 978-85.
PONTING, C. P., OLIVER, P. L. & REIK, W. 2009. Evolution and functions of long noncoding RNAs. *Cell,* 136**,** 629-41.
PRASANTH, K. V. & SPECTOR, D. L. 2007. Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum. *Genes Dev,* 21**,** 11-42.
PRICE, M. E., COTTON, A. M., LAM, L. L., FARRE, P., EMBERLY, E., BROWN, C. J., ROBINSON, W. P. & KOBOR, M. S. 2013. Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenetics Chromatin,* 6**,** 4.
PROVENCAL, N. & BINDER, E. B. 2015. The effects of early life stress on the epigenome: From the womb to adulthood and even before. *Exp Neurol,* 268**,** 10-20.
PROVENCAL, N., SUDERMAN, M. J., GUILLEMIN, C., MASSART, R., RUGGIERO, A., WANG, D., BENNETT, A. J., PIERRE, P. J., FRIEDMAN, D. P., COTE, S. M., HALLETT, M., TREMBLAY, R. E., SUOMI, S. J. & SZYF, M. 2012. The signature of maternal rearing in the methylome in rhesus macaque prefrontal cortex and T cells. *J Neurosci,* 32**,** 15626-42.
PROVENZI, L., GIORDA, R., BERI, S. & MONTIROSSO, R. 2016. SLC6A4 methylation as an epigenetic marker of life adversity exposures in humans: A systematic review of literature. *Neurosci Biobehav Rev,* 71**,** 7-20.
QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics,* 26**,** 841-2.
R CORE TEAM 2013. R: A language and environment for statistical computing.
RADA-IGLESIAS, A., BAJPAI, R., SWIGUT, T., BRUGMANN, S. A., FLYNN, R. A. & WYSOCKA, J. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature,* 470**,** 279-83.
RAKYAN, V. K., DOWN, T. A., BALDING, D. J. & BECK, S. 2011. Epigenome-wide association studies for common human diseases. *Nat Rev Genet,* 12**,** 529-41.
RANDO, O. J. 2016. Intergenerational Transfer of Epigenetic Information in Sperm. *Cold Spring Harb Perspect Med,* 6.
RAO, J. S., KELESHIAN, V. L., KLEIN, S. & RAPOPORT, S. I. 2012. Epigenetic modifications in frontal cortex from Alzheimer's disease and bipolar disorder patients. *Transl Psychiatry,* 2**,** e132.
REITZ, C., BRAYNE, C. & MAYEUX, R. 2011. Epidemiology of Alzheimer disease. *Nat Rev Neurol,* 7**,** 137-52.

RELTON, C. L. & DAVEY SMITH, G. 2012. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *Int J Epidemiol,* 41**,** 161-76.
RICHARDS, E. J. 2006. Inherited epigenetic variation--revisiting soft inheritance. *Nat Rev Genet,* 7**,** 395-401.
ROBERTSON, K. D. 2005. DNA methylation and human disease. *Nat Rev Genet,* 6**,** 597-610.
ROBINSON, M. D., MCCARTHY, D. J. & SMYTH, G. K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics,* 26**,** 139-40.
ROBINSON, P. N. & MUNDLOS, S. 2010. The human phenotype ontology. *Clin Genet,* 77**,** 525-34.
ROMENS, S. E., MCDONALD, J., SVAREN, J. & POLLAK, S. D. 2015. Associations between early life stress and gene methylation in children. *Child Dev,* 86**,** 303-9.
ROSENBAUM, P. R. 2002. Observational studies. *Observational Studies.* Springer.
ROTH, T. L., LUBIN, F. D., FUNK, A. J. & SWEATT, J. D. 2009. Lasting epigenetic influence of early-life adversity on the BDNF gene. *Biol Psychiatry,* 65**,** 760-9.
ROVELET-LECRUX, A., HANNEQUIN, D., RAUX, G., LE MEUR, N., LAQUERRIERE, A., VITAL, A., DUMANCHIN, C., FEUILLETTE, S., BRICE, A., VERCELLETTO, M., DUBAS, F., FREBOURG, T. & CAMPION, D. 2006. APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nat Genet,* 38**,** 24-6.
RUTTER, M. 1981. Stress, coping and development: some issues and some questions. *J Child Psychol Psychiatry,* 22**,** 323-56.
RUTTER, M., SONUGA-BARKE, E. J. & CASTLE, J. 2010. I. Investigating the impact of early institutional deprivation on development: background and research strategy of the English and Romanian Adoptees (ERA) study. *Monogr Soc Res Child Dev,* 75**,** 1-20.
SAHA, R. N. & PAHAN, K. 2006. HATs and HDACs in neurodegeneration: a tale of disconcerted acetylation homeostasis. *Cell Death Differ,* 13**,** 539-50.
SANTPERE, G., NIETO, M., PUIG, B. & FERRER, I. 2006. Abnormal Sp1 transcription factor expression in Alzheimer disease and tauopathies. *Neurosci Lett,* 397**,** 30-4.
SAUNDERSON, E. A., SPIERS, H., MIFSUD, K. R., GUTIERREZ-MECINAS, M., TROLLOPE, A. F., SHAIKH, A., MILL, J. & REUL, J. M. 2016. Stress-induced gene expression and behavior are controlled by DNA methylation and methyl donor availability in the dentate gyrus. *Proc Natl Acad Sci U S A,* 113**,** 4830-5.
SCHALKWYK, L. C., MEABURN, E. L., SMITH, R., DEMPSTER, E. L., JEFFRIES, A. R., DAVIES, M. N., PLOMIN, R. & MILL, J. 2010. Allelic skewing of DNA methylation is widespread across the genome. *Am J Hum Genet,* 86**,** 196-212.
SCHELLENBERG, G. D. & MONTINE, T. J. 2012. The genetics and neuropathology of Alzheimer's disease. *Acta Neuropathol,* 124**,** 305-23.
SCHEUNER, D., ECKMAN, C., JENSEN, M., SONG, X., CITRON, M., SUZUKI, N., BIRD, T. D., HARDY, J., HUTTON, M., KUKULL, W., LARSON, E., LEVY-LAHAD, E., VIITANEN, M., PESKIND, E., POORKAJ, P., SCHELLENBERG, G., TANZI, R., WASCO, W., LANNFELT, L., SELKOE, D. & YOUNKIN, S. 1996. Secreted amyloid beta-protein similar to that in the senile plaques of Alzheimer's disease is increased in vivo by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease. *Nat Med,* 2**,** 864-70.

SCHIZOPHRENIA WORKING GROUP OF THE PSYCHIATRIC GENOMICS CONSORTIUM 2014. Biological insights from 108 schizophrenia-associated genetic loci. *Nature,* 511**,** 421-7.

SCHMIDL, C., KLUG, M., BOELD, T. J., ANDREESEN, R., HOFFMANN, P., EDINGER, M. & REHLI, M. 2009. Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. *Genome Res,* 19**,** 1165-74.

SCHONES, D. E. & ZHAO, K. 2008. Genome-wide approaches to studying chromatin modifications. *Nat Rev Genet,* 9**,** 179-91.

SELKOE, D. J. & HARDY, J. 2016. The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Mol Med,* 8**,** 595-608.

SHARMA, S., KELLY, T. K. & JONES, P. A. 2010. Epigenetics in cancer. *Carcinogenesis,* 31**,** 27-36.

SHOEMAKER, R., DENG, J., WANG, W. & ZHANG, K. 2010. Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. *Genome Res,* 20**,** 883-9.

SHUKLA, S., KAVAK, E., GREGORY, M., IMASHIMIZU, M., SHUTINOSKI, B., KASHLEV, M., OBERDOERFFER, P., SANDBERG, R. & OBERDOERFFER, S. 2011. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature,* 479**,** 74-9.

SICKMUND, M. & PUZZANCHERA, C. 2014. Juvenile offenders and victims: 2014 national report.

SINGH, S. & LI, S. S. 2012. Epigenetic effects of environmental chemicals bisphenol A and phthalates. *Int J Mol Sci,* 13**,** 10143-53.

SPENCER, V. A. & DAVIE, J. R. 1999. Role of covalent modifications of histones in regulating gene expression. *Gene,* 240**,** 1-12.

SPIERS, H., HANNON, E., SCHALKWYK, L. C., SMITH, R., WONG, C. C., O'DONOVAN, M. C., BRAY, N. J. & MILL, J. 2015. Methylomic trajectories across human fetal brain development. *Genome Res,* 25**,** 338-52.

SPILLANTINI, M. G. & GOEDERT, M. 2013. Tau pathology and neurodegeneration. *Lancet Neurol,* 12**,** 609-22.

SPITZ, F. & FURLONG, E. E. 2012. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet,* 13**,** 613-26.

STEIGER, H., LABONTE, B., GROLEAU, P., TURECKI, G. & ISRAEL, M. 2013. Methylation of the glucocorticoid receptor gene promoter in bulimic women: associations with borderline personality disorder, suicidality, and exposure to childhood abuse. *Int J Eat Disord,* 46**,** 246-55.

STEWART, S. K., MORRIS, T. J., GUILHAMON, P., BULSTRODE, H., BACHMAN, M., BALASUBRAMANIAN, S. & BECK, S. 2015. oxBS-450K: a method for analysing hydroxymethylation using 450K BeadChips. *Methods,* 72**,** 9-15.

STRAHL, B. D. & ALLIS, C. D. 2000. The language of covalent histone modifications. *Nature,* 403**,** 41-5.

STROBEL, H. W., THOMPSON, C. M. & ANTONOVIC, L. 2001. Cytochromes P450 in brain: function and significance. *Curr Drug Metab,* 2**,** 199-214.

SUDERMAN, M., MCGOWAN, P. O., SASAKI, A., HUANG, T. C., HALLETT, M. T., MEANEY, M. J., TURECKI, G. & SZYF, M. 2012. Conserved epigenetic sensitivity to early life experience in the rat and human hippocampus. *Proc Natl Acad Sci U S A,* 109 Suppl 2**,** 17266-72.

SUN, W., POSCHMANN, J., CRUZ-HERRERA DEL ROSARIO, R., PARIKSHAK, N. N., HAJAN, H. S., KUMAR, V., RAMASAMY, R., BELGARD, T. G., ELANGGOVAN, B., WONG, C.

C., MILL, J., GESCHWIND, D. H. & PRABHAKAR, S. 2016. Histone Acetylome-wide Association Study of Autism Spectrum Disorder. *Cell,* 167**,** 1385-1397 e11.
SUN, X., HE, G., QING, H., ZHOU, W., DOBIE, F., CAI, F., STAUFENBIEL, M., HUANG, L. E. & SONG, W. 2006. Hypoxia facilitates Alzheimer's disease pathogenesis by up-regulating BACE1 gene expression. *Proc Natl Acad Sci U S A,* 103**,** 18727-32.
SZYF, M. 2015. Prospects for the development of epigenetic drugs for CNS conditions. *Nat Rev Drug Discov,* 14**,** 461-74.
SZYF, M. & BICK, J. 2013. DNA methylation: a mechanism for embedding early life experiences in the genome. *Child Dev,* 84**,** 49-57.
TAHILIANI, M., KOH, K. P., SHEN, Y., PASTOR, W. A., BANDUKWALA, H., BRUDNO, Y., AGARWAL, S., IYER, L. M., LIU, D. R., ARAVIND, L. & RAO, A. 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science,* 324**,** 930-5.
TANZI, R. E. 2012. The genetics of Alzheimer disease. *Cold Spring Harb Perspect Med,* 2.
TAYLOR, S. E. 2010. Mechanisms linking early life stress to adult health outcomes. *Proc Natl Acad Sci U S A,* 107**,** 8507-12.
TEICHER, M. H. & SAMSON, J. A. 2016. Annual Research Review: Enduring neurobiological effects of childhood abuse and neglect. *J Child Psychol Psychiatry,* 57**,** 241-66.
THOMAS-CHOLLIER, M., DARBO, E., HERRMANN, C., DEFRANCE, M., THIEFFRY, D. & VAN HELDEN, J. 2012a. A complete workflow for the analysis of full-size ChIP-seq (and similar) data sets using peak-motifs. *Nat Protoc,* 7**,** 1551-68.
THOMAS-CHOLLIER, M., HERRMANN, C., DEFRANCE, M., SAND, O., THIEFFRY, D. & VAN HELDEN, J. 2012b. RSAT peak-motifs: motif analysis in full-size ChIP-seq datasets. *Nucleic Acids Res,* 40**,** e31.
THOMPSON, R. F., ATZMON, G., GHEORGHE, C., LIANG, H. Q., LOWES, C., GREALLY, J. M. & BARZILAI, N. 2010. Tissue-specific dysregulation of DNA methylation in aging. *Aging Cell,* 9**,** 506-18.
TOBI, E. W., LUMEY, L. H., TALENS, R. P., KREMER, D., PUTTER, H., STEIN, A. D., SLAGBOOM, P. E. & HEIJMANS, B. T. 2009. DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. *Hum Mol Genet,* 18**,** 4046-53.
TODKAR, A., GRANHOLM, L., ALJUMAH, M., NILSSON, K. W., COMASCO, E. & NYLANDER, I. 2015. HPA Axis Gene Expression and DNA Methylation Profiles in Rats Exposed to Early Life Stress, Adult Voluntary Ethanol Drinking and Single Housing. *Front Mol Neurosci,* 8**,** 90.
TROUTON, A., SPINATH, F. M. & PLOMIN, R. 2002. Twins early development study (TEDS): a multivariate, longitudinal genetic investigation of language, cognition and behavior problems in childhood. *Twin Res,* 5**,** 444-8.
TRZESNIEWSKI, K. H., MOFFITT, T. E., CASPI, A., TAYLOR, A. & MAUGHAN, B. 2006. Revisiting the association between reading achievement and antisocial behavior: new evidence of an environmental explanation from a twin study. *Child Dev,* 77**,** 72-88.
TSAPROUNI, L. G., YANG, T. P., BELL, J., DICK, K. J., KANONI, S., NISBET, J., VINUELA, A., GRUNDBERG, E., NELSON, C. P., MEDURI, E., BUIL, A., CAMBIEN, F., HENGSTENBERG, C., ERDMANN, J., SCHUNKERT, H., GOODALL, A. H., OUWEHAND, W. H., DERMITZAKIS, E., SPECTOR, T. D., SAMANI, N. J. & DELOUKAS, P. 2014. Cigarette smoking reduces DNA methylation levels at

multiple genomic loci but the effect is partially reversible upon cessation. *Epigenetics,* 9**,** 1382-96.
TYCKO, B. 2010. Allele-specific DNA methylation: beyond imprinting. *Hum Mol Genet,* 19**,** R210-20.
TYRKA, A. R., PRICE, L. H., MARSIT, C., WALTERS, O. C. & CARPENTER, L. L. 2012. Childhood adversity and epigenetic modulation of the leukocyte glucocorticoid receptor: preliminary findings in healthy adults. *PLoS One,* 7**,** e30148.
TYRKA, A. R., RIDOUT, K. K., PARADE, S. H., PAQUETTE, A., MARSIT, C. J. & SEIFER, R. 2015. Childhood maltreatment and methylation of FK506 binding protein 5 gene (FKBP5). *Dev Psychopathol,* 27**,** 1637-45.
UDDIN, M., KOENEN, K. C., AIELLO, A. E., WILDMAN, D. E., DE LOS SANTOS, R. & GALEA, S. 2011. Epigenetic and inflammatory marker profiles associated with depression in a community-based epidemiologic sample. *Psychol Med,* 41**,** 997-1007.
UNTERNAEHRER, E., LUERS, P., MILL, J., DEMPSTER, E., MEYER, A. H., STAEHLI, S., LIEB, R., HELLHAMMER, D. H. & MEINLSCHMIDT, G. 2012. Dynamic changes in DNA methylation of stress-associated genes (OXTR, BDNF) after acute psychosocial stress. *Transl Psychiatry,* 2**,** e150.
VAN DER KNAAP, L. J., RIESE, H., HUDZIAK, J. J., VERBIEST, M. M., VERHULST, F. C., OLDEHINKEL, A. J. & VAN OORT, F. V. 2014. Glucocorticoid receptor gene (NR3C1) methylation following stressful events between birth and adolescence. The TRAILS study. *Transl Psychiatry,* 4**,** e381.
VAN DONGEN, J., SLAGBOOM, P. E., DRAISMA, H. H., MARTIN, N. G. & BOOMSMA, D. I. 2012. The continuing value of twin studies in the omics era. *Nat Rev Genet,* 13**,** 640-53.
VENKATESH, S. & WORKMAN, J. L. 2015. Histone exchange, chromatin structure and the regulation of transcription. *Nat Rev Mol Cell Biol,* 16**,** 178-89.
VILLEGAS-LLERENA, C., PHILLIPS, A., GARCIA-REITBOECK, P., HARDY, J. & POCOCK, J. M. 2016. Microglial genes regulating neuroinflammation in the progression of Alzheimer's disease. *Curr Opin Neurobiol,* 36**,** 74-81.
VOLPE, T. A., KIDNER, C., HALL, I. M., TENG, G., GREWAL, S. I. & MARTIENSSEN, R. A. 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science,* 297**,** 1833-7.
WADDINGTON, C. H. 2014. *The strategy of the genes*, Routledge.
WADE, P. A. & WOLFFE, A. P. 2001. ReCoGnizing methylated DNA. *Nat Struct Biol,* 8**,** 575-7.
WAGNER, J. R., BUSCHE, S., GE, B., KWAN, T., PASTINEN, T. & BLANCHETTE, M. 2014. The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. *Genome Biol,* 15**,** R37.
WARREN, J. D., ROHRER, J. D. & ROSSOR, M. N. 2013. Frontotemporal dementia. *Bmj,* 347**,** f4827.
WATT, F. & MOLLOY, P. L. 1988. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev,* 2**,** 1136-43.
WEAVER, I. C., CERVONI, N., CHAMPAGNE, F. A., D'ALESSIO, A. C., SHARMA, S., SECKL, J. R., DYMOV, S., SZYF, M. & MEANEY, M. J. 2004. Epigenetic programming by maternal behavior. *Nat Neurosci,* 7**,** 847-54.
WEBER, M., DAVIES, J. J., WITTIG, D., OAKELEY, E. J., HAASE, M., LAM, W. L. & SCHUBELER, D. 2005. Chromosome-wide and promoter-specific analyses

identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet,* 37**,** 853-62.

WEDER, N., ZHANG, H., JENSEN, K., YANG, B. Z., SIMEN, A., JACKOWSKI, A., LIPSCHITZ, D., DOUGLAS-PALUMBERI, H., GE, M., PEREPLETCHIKOVA, F., O'LOUGHLIN, K., HUDZIAK, J. J., GELERNTER, J. & KAUFMAN, J. 2014. Child abuse, depression, and methylation in genes involved with stress, neural plasticity, and brain circuitry. *J Am Acad Child Adolesc Psychiatry,* 53**,** 417-24 e5.

WENK, G. L. 2003. Neuropathologic changes in Alzheimer's disease. *J Clin Psychiatry,* 64 Suppl 9**,** 7-10.

WHITEFORD, H. A., DEGENHARDT, L., REHM, J., BAXTER, A. J., FERRARI, A. J., ERSKINE, H. E., CHARLSON, F. J., NORMAN, R. E., FLAXMAN, A. D., JOHNS, N., BURSTEIN, R., MURRAY, C. J. & VOS, T. 2013. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet,* 382**,** 1575-86.

WIDOM, C. S., CZAJA, S. J., BENTLEY, T. & JOHNSON, M. S. 2012. A prospective investigation of physical health outcomes in abused and neglected children: new findings from a 30-year follow-up. *Am J Public Health,* 102**,** 1135-44.

WIDOM, C. S., CZAJA, S. J. & DUTTON, M. A. 2008. Childhood victimization and lifetime revictimization. *Child Abuse Negl,* 32**,** 785-96.

WILLEMS, T., GYMREK, M., HIGHNAM, G., GENOMES PROJECT, C., MITTELMAN, D. & ERLICH, Y. 2014. The landscape of human STR variation. *Genome Res,* 24**,** 1894-904.

WILMOT, B., FRY, R., SMEESTER, L., MUSSER, E. D., MILL, J. & NIGG, J. T. 2016. Methylomic analysis of salivary DNA in childhood ADHD identifies altered DNA methylation in VIPR2. *J Child Psychol Psychiatry,* 57**,** 152-60.

WOLF, E. J., LOGUE, M. W., HAYES, J. P., SADEH, N., SCHICHMAN, S. A., STONE, A., SALAT, D. H., MILBERG, W., MCGLINCHEY, R. & MILLER, M. W. 2016. Accelerated DNA methylation age: Associations with PTSD and neural integrity. *Psychoneuroendocrinology,* 63**,** 155-62.

WONG, C. C., CASPI, A., WILLIAMS, B., CRAIG, I. W., HOUTS, R., AMBLER, A., MOFFITT, T. E. & MILL, J. 2010. A longitudinal study of epigenetic variation in twins. *Epigenetics,* 5**,** 516-26.

WOODCOCK, C. L. & DIMITROV, S. 2001. Higher-order structure of chromatin and chromosomes. *Curr Opin Genet Dev,* 11**,** 130-5.

WU, H. & SUN, Y. E. 2006. Epigenetic regulation of stem cell differentiation. *Pediatr Res,* 59**,** 21R-5R.

WU, S. C. & ZHANG, Y. 2010. Active DNA demethylation: many roads lead to Rome. *Nat Rev Mol Cell Biol,* 11**,** 607-20.

XIAO, Y., CAMARILLO, C., PING, Y., ARANA, T. B., ZHAO, H., THOMPSON, P. M., XU, C., SU, B. B., FAN, H., ORDONEZ, J., WANG, L., MAO, C., ZHANG, Y., CRUZ, D., ESCAMILLA, M. A., LI, X. & XU, C. 2014. The DNA methylome and transcriptome of different brain regions in schizophrenia and bipolar disorder. *PLoS One,* 9**,** e95875.

XU, G. L., BESTOR, T. H., BOURC'HIS, D., HSIEH, C. L., TOMMERUP, N., BUGGE, M., HULTEN, M., QU, X., RUSSO, J. J. & VIEGAS-PEQUIGNOT, E. 1999. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature,* 402**,** 187-91.

XU, Y., CHEN, X. T., LUO, M., TANG, Y., ZHANG, G., WU, D., YANG, B., RUAN, D. Y. & WANG, H. L. 2015. Multiple epigenetic factors predict the attention

deficit/hyperactivity disorder among the Chinese Han children. *J Psychiatr Res,* 64**,** 40-50.

YAUK, C., POLYZOS, A., ROWAN-CARROLL, A., SOMERS, C. M., GODSCHALK, R. W., VAN SCHOOTEN, F. J., BERNDT, M. L., POGRIBNY, I. P., KOTURBASH, I., WILLIAMS, A., DOUGLAS, G. R. & KOVALCHUK, O. 2008. Germ-line mutations, DNA damage, and global hypermethylation in mice exposed to particulate air pollution in an urban/industrial location. *Proc Natl Acad Sci U S A,* 105**,** 605-10.

YEHUDA, R., DASKALAKIS, N. P., BIERER, L. M., BADER, H. N., KLENGEL, T., HOLSBOER, F. & BINDER, E. B. 2016. Holocaust Exposure Induced Intergenerational Effects on FKBP5 Methylation. *Biol Psychiatry,* 80**,** 372-80.

YIN, T., COOK, D. & LAWRENCE, M. 2012. ggbio: an R package for extending the grammar of graphics for genomic data. *Genome Biol,* 13**,** R77.

YODER, J. A., WALSH, C. P. & BESTOR, T. H. 1997. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet,* 13**,** 335-40.

ZANNAS, A. S., ARLOTH, J., CARRILLO-ROA, T., IURATO, S., ROH, S., RESSLER, K. J., NEMEROFF, C. B., SMITH, A. K., BRADLEY, B., HEIM, C., MENKE, A., LANGE, J. F., BRUCKL, T., ISING, M., WRAY, N. R., ERHARDT, A., BINDER, E. B. & MEHTA, D. 2015. Lifetime stress accelerates epigenetic aging in an urban, African American cohort: relevance of glucocorticoid signaling. *Genome Biol,* 16**,** 266.

ZARREI, M., MACDONALD, J. R., MERICO, D. & SCHERER, S. W. 2015. A copy number variation map of the human genome. *Nat Rev Genet,* 16**,** 172-83.

ZEILEIS, A. 2004. Econometric Computing with HC and HAC Covariance Matrix Estimators. *Journal of Statistical Software,* 11**,** 1-17.

ZEILINGER, S., KUHNEL, B., KLOPP, N., BAURECHT, H., KLEINSCHMIDT, A., GIEGER, C., WEIDINGER, S., LATTKA, E., ADAMSKI, J., PETERS, A., STRAUCH, K., WALDENBERGER, M. & ILLIG, T. 2013. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS One,* 8**,** e63812.

ZHANG, D., CHENG, L., BADNER, J. A., CHEN, C., CHEN, Q., LUO, W., CRAIG, D. W., REDMAN, M., GERSHON, E. S. & LIU, C. 2010. Genetic control of individual differences in gene-specific methylation in human brain. *Am J Hum Genet,* 86**,** 411-9.

ZHANG, K., SCHRAG, M., CROFTON, A., TRIVEDI, R., VINTERS, H. & KIRSCH, W. 2012. Targeted proteomics for quantification of histone acetylation in Alzheimer's disease. *Proteomics,* 12**,** 1261-8.

ZHANG, Y., LIU, T., MEYER, C. A., EECKHOUTE, J., JOHNSON, D. S., BERNSTEIN, B. E., NUSBAUM, C., MYERS, R. M., BROWN, M., LI, W. & LIU, X. S. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol,* 9**,** R137.

ZHU, L., WANG, X., LI, X. L., TOWERS, A., CAO, X., WANG, P., BOWMAN, R., YANG, H., GOLDSTEIN, J., LI, Y. J. & JIANG, Y. H. 2014. Epigenetic dysregulation of SHANK3 in brain tissues from individuals with autism spectrum disorders. *Hum Mol Genet,* 23**,** 1563-78.

ZILLER, M. J., EDRI, R., YAFFE, Y., DONAGHEY, J., POP, R., MALLARD, W., ISSNER, R., GIFFORD, C. A., GOREN, A., XING, J., GU, H., CACCHIARELLI, D., TSANKOV, A. M., EPSTEIN, C., RINN, J. L., MIKKELSEN, T. S., KOHLBACHER, O., GNIRKE, A., BERNSTEIN, B. E., ELKABETZ, Y. & MEISSNER, A. 2015. Dissecting neural differentiation regulatory networks through epigenetic footprinting. *Nature,* 518**,** 355-9.

ZLOKOVIC, B. V. 2011. Neurovascular pathways to neurodegeneration in Alzheimer's disease and other disorders. *Nat Rev Neurosci,* 12**,** 723-38.

Appendices (on CD-ROM)

Appendix A

Supplementary file 1: Association statistics for all probes annotated to the seven candidate genes – *AVP*, *BDNF*, *CRHR1*, *CYP2E1*, *FKBP5*, *NR3C1*, *SLC6A4* – examined in **Chapter 3**.

Appendix B

Supplementary file 1: Association statistics for the 1,475 hyperacetylated peaks (FDR < 0.05) identified in **Chapter 5**.

Supplementary file 2: Association statistics for the 2,687 hypoacetylated peaks (FDR < 0.05) identified in **Chapter 5**.

Appendix C

Supplementary file 1: Gene ontology enrichments of hyperacetylated peaks (FDR < 0.05) in the category “Biological Process” (**Chapter 5**)

Supplementary file 2: Gene ontology enrichments of hyperacetylated peaks (FDR < 0.05) in the category “Molecular Function” (**Chapter 5**)

Supplementary file 3: Gene ontology enrichments of hyperacetylated peaks (FDR < 0.05) in the category “Disease Ontology” (**Chapter 5**)

Supplementary file 4: Gene ontology enrichments of hyperacetylated peaks (FDR < 0.05) in the category “Human Phenotype” (**Chapter 5**)

Supplementary file 5: Gene ontology enrichments of hypoacetylated peaks (FDR < 0.05) in the category “Biological Process” (**Chapter 5**)

Supplementary file 6: Gene ontology enrichments of hypoacetylated peaks (FDR < 0.05) in the category “Molecular Function” (**Chapter 5**)

Supplementary file 7: Gene ontology enrichments of hypoacetylated peaks (FDR < 0.05) in the category “Disease Ontology” (**Chapter 5**)

Supplementary file 8: Gene ontology enrichments of hypoacetylated peaks (FDR < 0.05) in the category “Human Phenotype” (**Chapter 5**)